

Universidade Federal do Rio de Janeiro

Escola Politécnica

Departamento de Eletrônica e de Computação

Modelagem de Prosódia para Conversores Texto-Fala

Autor:

Lucas Arrabal Thomaz

Orientador:

Prof. Sergio Lima Netto, Ph. D.

Examinador:

Prof. Luiz Wagner Pereira Biscainho, D.Sc.

Examinador:

Vagner Luis Latsch, D.Sc.

DEL

Setembro de 2012

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

Escola Politécnica - Departamento de Eletrônica e de Computação

Centro de Tecnologia, bloco H, sala H-217, Cidade Universitária

Rio de Janeiro - RJ CEP 21949-900

Este exemplar é de propriedade da Universidade Federal do Rio de Janeiro, que poderá incluí-lo em base de dados, armazenar em computador, microfilmear ou adotar qualquer forma de arquivamento.

É permitida a menção, reprodução parcial ou integral e a transmissão entre bibliotecas deste trabalho, sem modificação de seu texto, em qualquer meio que esteja ou venha a ser fixado, para pesquisa acadêmica, comentários e citações, desde que sem finalidade comercial e que seja feita a referência bibliográfica completa.

Os conceitos expressos neste trabalho são de responsabilidade do(s) autor(es) e do(s) orientador(es).

“-Escute só - disse Ford, ainda fascinado pelo folheto - Diz coisas fantásticas sobre a cibernética da nave. *Uma nova geração de robôs e computadores da Companhia Cibernética de Sirius, contando com o novo recurso PGH.*

-O que é PGH? - Disse Arthur?

-Diz que é ‘Personalidade Humana Genuína’.

-Que coisa horrível - disse Arthur

-Põe horrível nisso - disse uma voz atrás deles. (...) - É horrível - prosseguiu Marvin - Tudo isso. Medonho. Melhor nem falar nisso. Vejam essa porta - disse, entrando. Os circuitos de ironia começaram a atuar sobre seu modulador de voz, e

Marvin pôs-se a parodiar o estilo do folheto de vendas.

-Todas as portas dessa nave são alegres e bem humoradas. É um prazer para elas abrir para você, e fechar de novo com a consciência de quem fez um serviço bem-feito.

Ao fechar-se, a porta realmente parecia dar um suspiro de satisfação:

‘Hummmmmmmmmmmmmmmmmmmmmmmmm ah!’”

(Trecho do Livro “O Guia do Mochileiro das Galáxias” de Douglas Adams)

AGRADECIMENTO

Agradeço inicialmente a minha família, pela formação pessoal e educacional que me proporcionou ao longo dos anos. Pelo apoio e incentivo em todos os momentos do meu caminho, antes e durante o curso de graduação, me motivando e cobrando quando necessário e sempre estando disponíveis quando precisei de alguma ajuda.

Agradeço também aos meus amigos e colegas, que durante todo esse percurso caminharam ao meu lado proporcionando momentos de diversão e colaborando para que esse projeto fosse concluído. Agradeço a todos que me acompanharem ao longo desses anos.

Agradeço ainda, ao meu orientador, professor Sérgio Lima Netto, que acompanhou minha jornada nos últimos anos e me indicou os caminhos a seguir, culminando nesse trabalho. Agradeço por ter confiado em mim para que chegasse até aqui, por expandir meus horizontes para áreas que eu não conhecia e por abrir as portas para novas oportunidades.

Por fim, agradeço a todos que de alguma maneira se fizeram presentes e colaboraram para a conclusão desse ciclo.

RESUMO

Para que um conversor texto-fala (TTS) produza uma saída que se aproxime da fala humana é necessário que lhe seja incorporada uma etapa de manipulação prosódica de atitudes, a fim de garantir a sintetização adequada das características pragmáticas do seu conteúdo.

LATSCH [1] criou o sistema SASPRO (Sistema de Análise e Síntese da Prosódia), que implementa um conversor TTS não comercial, voltado para o uso com o português brasileiro. Por não ser exatamente um sistema TTS completo, uma de suas limitações está relacionada à estrutura da sentença que deve ser utilizada de modo a se obter, como sinal sintetizado, um sinal cujas características prosódicas estejam de acordo com a atitude esperada.

Este trabalho tem como objetivo a ampliação da funcionalidade de manipulação prosódica das atitudes do sistema SASPRO, permitindo que opere com uma maior variedade de estruturas silábicas. O produto final é um algoritmo que faz o transplante de prosódia de uma atitude para uma sentença, que pode conter uma estrutura silábica menos limitada que a versão anterior do sistema SASPRO. Como resultado do trabalho, 9 das 14 atitudes modeladas foram avaliadas como aceitáveis, de forma que o trabalho alcançou o seu objetivo de implementar um método de imposição de atitudes que independa da estrutura silábica.

Palavras-Chave: conversor texto-fala, manipulação prosódica, atitudes, transplante paramétrico de prosódia.

ABSTRACT

For a text-to-speech (TTS) conversor to produce a signal that is similar to the human speech it needs to have a prosodic manipulation step in it, since that is the only way the system can synthesize the context characteristics of the sentences.

In his doctor thesis LATSCH [1] creates the SASPRO system, that implements a non-commercial TTS conversor, focused in the Brazilian Portuguese idiom. Since it is not a complete TTS system, there are some limitations like the syllabic structure of the sentence that you have to use to have the prosodic characteristics modified according to the attitude of interest.

The focus of this work is the expansion of the prosodic manipulation functionality of the SASPRO System, making it independent of the syllabic structure of the sentence. The goal to achieve here is an algorithm that implements the prosody transplant of an attitude to a sentence with less limitations to the syllabic structure than the previous stage of the SASPRO system. As a result of this project 9 of the 14 modeled attitudes were evaluated as acceptable, so the project achieved its goal of implementing a method of attitude imposition independent of the syllabic structure.

Key-words: text-to-speech, prosodic manipulation, attitudes, parametric transplant of prosody.

SIGLAS

UFRJ - Universidade Federal do Rio de Janeiro

DEL - Departamento de Engenharia Eletrônica e de Computação

SASPRO - Sistema de Análise e Síntese da Prosódia

GPS - Global Positioning System

OCR - Optical Character Recognition

PSOLA - Pitch Synchronous Overlap and Add

SFC - Superposition of Functional Contours

MOS - Mean Opinion Score

Sumário

1	Introdução	1
1.1	Tema	1
1.2	Justificativa	1
1.3	Aplicações	3
1.4	Objetivos	3
1.5	Metodologia	4
1.6	Descrição	4
2	Classificação e Transplante de Atitudes	6
2.1	Introdução	6
2.2	Caracterização das Atitudes	7
2.3	Leitura dos Gráficos	9
2.4	Caracterização Prosódica das Atitudes	11
2.4.1	Atitudes com queda de <i>pitch</i>	11
2.4.2	Atitudes com subida de pitch	14
2.4.3	Atitudes com variação de duração	16
2.5	Transplante de prosódia por cópia	19
2.6	Transplante de prosódia por superposição	21
2.7	Conclusões	23
3	Classificação de Atitudes em Estruturas Silábicas Diferentes	25
3.1	Introdução	25
3.2	Caracterização Prosódica das Atitudes em Estrutura Silábica Alternativa	26
3.2.1	Sinais com queda de <i>pitch</i>	26

3.2.2	Sinais com subida de pitch	29
3.2.3	Sinais com variação de duração	31
3.3	Análise da Validade do Modelo	33
3.3.1	Atitudes com queda de <i>pitch</i>	34
3.3.2	Atitudes com subida de pitch	36
3.3.3	Atitudes com variação de duração	37
3.4	Conclusões	38
4	Aplicação do Modelo e Testes	40
4.1	Introdução	40
4.2	Aplicação do Modelo Estudado	40
4.2.1	A nova estrutura silábica	41
4.2.2	Imposição do modelo	42
4.3	Testes Aplicados	46
4.4	Resultados e Análises	50
4.5	Conclusões	54
5	Conclusões	55
5.1	Considerações Finais	55
5.2	Trabalhos Futuros	57
	Bibliografia	58

Lista de Figuras

2.1	Exemplo de representação gráfica das características prosódicas das atitudes.	10
2.2	Características prosódicas da sentença “Renata Jogava.” com atitude Óbvia.	12
2.3	Características prosódicas da sentença “Repete denovo.” com atitude Sugestão.	12
2.4	Características prosódicas da sentença “Renata Jogava.” com atitude Sugestão.	13
2.5	Características prosódicas da sentença “E como jogava.” com atitude Sugestão.	13
2.6	Características prosódicas da sentença “Repete denovo.” com atitude Sugestão.	14
2.7	Características prosódicas da sentença “E como jogava.” com atitude Sugestão.	14
2.8	Características prosódicas da sentença “Renata Jogava.” com atitude Sugestão.	15
2.9	Características prosódicas da sentença “Renata Jogava.” com atitude Sugestão.	15
2.10	Características prosódicas da sentença “Repete denovo.” com atitude Sugestão.	16
2.11	Características prosódicas da sentença “Renata Jogava.” com atitude Questão SN Incrédula.	17
2.12	Características prosódicas da sentença “Renata Jogava.” com atitude Sugestão.	17
2.13	Características prosódicas da sentença “Renata Jogava.” com atitude Sugestão.	18

2.14	Características prosódicas da sentença “Renata Jogava.” com atitude Sugestão.	18
2.15	Características prosódicas da sentença “Renata Jogava.” com atitude Sugestão.	19
2.16	Esquemático do transplante por cópia.	21
2.17	Exemplo do transplante por cópia.	22
2.18	Esquemático do transplante por superposição.	23
3.1	Características prosódicas da sentença “Renata Jogava.” com atitude Óbvia.	27
3.2	Características prosódicas da sentença “Cala-te Jogador.” com atitude Sugestão.	27
3.3	Características prosódicas da sentença “Pífanos Tocarão.” com atitude Ênfase Contrastiva.	27
3.4	Características prosódicas da sentença “Quando que Passará.” com atitude Questão QU.	28
3.5	Características prosódicas da sentença “Cala-te Jogador.” com atitude Comando.	28
3.6	Características prosódicas da sentença “Quando que Passará.” com atitude Exclamação QU.	29
3.7	Características prosódicas da sentença “Pífanos Tocarão.” com atitude Questão SN.	30
3.8	Características prosódicas da sentença “Pífanos Tocarão.” com atitude Questão SN Retórica.	30
3.9	Características prosódicas da sentença “Cala-te Jogador.” com atitude Pedido.	31
3.10	Características prosódicas da sentença “Pífanos Tocarão.” com atitude Questão SN Incrédula.	31
3.11	Características prosódicas da sentença “Pífanos Tocarão.” com atitude Aviso.	32
3.12	Características prosódicas da sentença “Pífanos Tocarão.” com atitude Irônica.	32
3.13	Características prosódicas da sentença “Pífanos Tocarão.” com atitude Incrédula.	33
3.14	Características prosódicas da sentença “Pífanos Tocarão.” com atitude Ênfase Intensiva.	33

4.1	Características prosódicas da sentença “Renata jogava.” com atitude Irônica.	43
4.2	Janela de Edição de Duração do Prosody Editor	43
4.3	Janela de Edição de Intensidade do Prosody Editor	44
4.4	Janela de Edição de <i>Pitch</i> do Praat - Sentença de Referência	45
4.5	Janela de Edição de <i>Pitch</i> do Praat - Sentença Manipulada	46
4.6	Tela inicial do programa de avaliação subjetiva de atitudes	48
4.7	Tela de avaliação do programa de avaliação subjetiva de atitudes	49
4.8	Gráficos com as médias dos testes subjetivos aplicados	50

Lista de Tabelas

4.1	Médias dos testes subjetivos aplicados	50
4.2	Respostas dadas para a pergunta 1: “Qual o par de sinais para o qual encontrou mais dificuldades para atribuir nota?”.	52
4.3	Respostas dadas para a pergunta 2: “Houve alguma dúvida quanto ao procedimento de avaliação?”.	53
4.4	Respostas dadas para a pergunta 3: “Houve alguma atitude que não foi capaz de identificar?”.	53

Capítulo 1

Introdução

1.1 Tema

Um sistema de conversão Texto-Fala (TTS, do inglês *Text-To-Speech*) tem a função de converter uma sequência de caracteres que formam uma sentença num dado idioma para um sinal de voz que transmita a mesma mensagem, de forma clara e de fácil entendimento. Um sistema TTS será considerado tão melhor quão mais próximo ele se tornar de uma fala humana natural. Dentre as muitas melhorias que um sistema desse tipo pode ter para que se pareça cada vez mais com a fala humana destaca-se a modelagem de prosódia (LATSCH [1]).

A modelagem de prosódia num sistema TTS tem por objetivo capacitar o sistema para imitar as diferentes atitudes (como por exemplo obviedade, ironia, interrogação) e emoções que são expressas por um ser humano em seu discurso, mas que muitas vezes podem passar despercebidas na modalidade escrita da língua. Dessa forma esse projeto tem como campo de estudo essa área específica do processamento e síntese de sinais de voz.

1.2 Justificativa

Os sistemas de conversão texto-fala são, atualmente, amplamente utilizados no cotidiano, muitas vezes passando despercebidos pelas pessoas que os utilizam. Alguns sistemas que frequentemente utilizam esse tipo de conversão são, por exemplo,

navegadores GPS (*Global Positioning System*), sistemas de tradução e ensino de idiomas e ferramentas de apoio a deficientes visuais (que muitas vezes utilizam o TTS associado a um sistema OCR - *optical character recognition*), entre outros.

Por ser a voz o sistema de comunicação natural da maioria da população, sua importância como interface principal de acesso a dispositivos (como computadores, celulares e até mesmo eletrodomésticos) tende somente a crescer nos próximos anos. Entretanto, para que o uso desses sistemas cresça de maneira significativa e se dê de forma mais natural, alguns aspectos de sua tecnologia ainda devem ser melhorados. Um dos pontos que devem ser trabalhados é a modelagem prosódica das atitudes de uma sentença, já que sem essa modelagem os sistemas se tornam artificiais, criando sinais de voz sintéticos que são incapazes de se passar por seus equivalentes naturais.

Considerando que uma língua é composta por diversos níveis, sendo eles fonológico, morfológico, sintático, semântico e pragmático, um sistema TTS que não possua uma etapa de manipulação prosódica poderá não ser capaz de sintetizar adequadamente os últimos níveis, perdendo, dessa forma, boa parte da informação que a sentença deveria transmitir, conforme dito por DUTOIT e STYLIANOU[2] e SIMÕES [3]. De maneira a permitir que o sistema seja capaz de apresentar as características dos níveis sintático e semântico, sem que haja perdas consideráveis de compreensão, será necessário ainda que o sistema inclua uma manipulação prosódica que enfatize as palavras principais da sentença.

De uma maneira ainda mais profunda, para que o sistema seja capaz de representar completamente as características de um orador humano, e contemplar todos os níveis linguísticos de uma sentença, este deve ser capaz de trabalhar com as características prosódicas que definem uma atitude, situando a sentença sintetizada no contexto em que se encontra. Isto facilita o entendimento por parte do ouvinte e cumpre o requisito de trabalhar adequadamente o nível pragmático da língua.

Para se construir um sistema TTS é necessário que todo o seu trabalho seja focado diretamente no idioma para o qual ele será implementado. Especificamente, a modelagem das atitudes a serem implementadas no sistema deve ser feita segundo

as características desse idioma.

1.3 Aplicações

Uma vez concluído o algoritmo de manipulação prosódica de atitudes, o mesmo seria incorporado a um sistema de conversão texto-fala. Nesse sistema o algoritmo seria responsável por imbuir o sinal, já sintetizado pelo TTS, de uma atitude previamente especificada, já que o sistema não seria capaz de interpretar pelo contexto a atitude que o usuário deseja passar com aquela sentença. O trecho da sentença que deverá receber essa atitude também deve ser informado ao sistema para que, a partir dessa informações, o mesmo seja capaz de aplicar o algoritmo.

Algo que deve ser levado em conta na implementação de um sistema com a manipulação prosódica de atitudes é a presença de regionalismos e sotaques nas atitudes. Uma vez que as características prosódicas das atitudes dependem desses regionalismos, é importante notar que a manipulação das atitudes será feita de acordo com as características presentes no modelo de cada atitude, que é treinado com as características de uma certa região.

Ainda deve ser observado que, uma vez que os fonemas que compõem as sentenças geradas pelo TTS são gravadas por um mesmo locutor, para que o som seja sintetizado de maneira mais natural, o mesmo deve ser feito com as sentenças cujas atitudes serão utilizadas como referência.

1.4 Objetivos

O objetivo desse trabalho é o desenvolvimento de um algoritmo de manipulação prosódica de atitudes que permita o transplante de atitudes entre sentenças com diferentes estruturas silábicas. Com isto será possível para o sistema SASPRO em LATSCH [1] manipular um número maior de sentenças, evitando um banco de sentenças com cada atitude, o que seria inviável do ponto de vista prático

Ao longo desse trabalho será utilizado o sistema SASPRO, cuja manipulação prosódica de atitudes foi desenvolvida seguindo o modelo de MORAES [4], voltado diretamente para o português brasileiro. O sistema SASPRO, embora não seja um sistema comercial de TTS, por não ser capaz de sintetizar todas as palavras presentes no nosso idioma e possuir um banco de fonemas limitado, é capaz de sintetizar sinais de voz a partir de texto e utilizar algoritmos de manipulação prosódica para realizar a transformação entre diferentes atitudes. Entretanto, sua manipulação está restrita a uma dada estrutura da sentença; o objetivo desse trabalho é ampliar o escopo desse sistema, permitindo seu uso numa gama maior de sentenças.

1.5 Metodologia

Para a execução desse trabalho serão cumpridas as seguintes etapas:

- Definição e caracterização das atitudes aqui consideradas;
- Estudo dos métodos utilizados atualmente no sistema para a manipulação das atitudes;
- Estudo das características prosódicas relacionadas a cada uma das atitudes que serão trabalhadas nesse projeto;
- Validação das estruturas que caracterizam as atitudes para estruturas silábicas diferentes daquelas para as quais foram projetadas;
- Desenvolvimento do algoritmo de transplante de prosódia que independa da estrutura silábica da sentença;
- Desenvolvimento e aplicação de um teste subjetivo para a validação do método desenvolvido.

1.6 Descrição

Esse trabalho seguirá a estrutura de capítulos descrita a seguir:

No Capítulo 2 serão abordadas uma revisão bibliográfica das técnicas utilizadas bem como uma análise das limitações encontradas no sistema SASPRO, de modo a melhor compreender as modificações que devem ser executadas no algoritmo de manipulação para que os resultados sejam alcançados adequadamente. Também nesse capítulo serão explicadas as características que formam cada uma das atitudes contempladas aqui e os seus usos.

O Capítulo 3 é composto por uma verificação da validade dos modelos propostos no Capítulo 2 para caracterizar as atitudes aplicados a uma outra estrutura silábica, que ainda não havia sido testada. Isto permite que o modelo seja estendido para outras estruturas no capítulo seguinte.

A aplicação dos modelos de prosódia sobre sentenças de atitude neutra será abordada no Capítulo 4. Nele serão mostradas as etapas de seu desenvolvimento bem como seu funcionamento. Além disso, será apresentado o teste subjetivo usado para validar o funcionamento do algoritmo.

Capítulo 2

Classificação e Transplante de Atitudes

2.1 Introdução

De maneira a melhor compreender o problema da modelagem prosódica de atitudes e do transplante de atitudes entre sentenças distintas, serão apresentadas, nesse capítulo, técnicas de transplante, além de uma análise e classificação das ditas atitudes, conforme foram utilizadas em LATSCH [1].

Na Seção 2.2 serão apresentadas as atitudes utilizadas ao longo deste trabalho bem como suas descrições intuitivas e alguns casos de uso.

Dando prosseguimento ao capítulo, será vista na Seção 2.4 a descrição das diferentes atitudes conforme suas características prosódicas e seus usos por um orador, de forma que se possa entender melhor suas diferenças.

Na Seção 2.5 será apresentada a técnica de transplante de atitudes por cópia, na qual uma sentença com atitude neutra tem seus parâmetros modificados para que sejam iguais aos de outra sentença de mesma estrutura silábica, mas com a atitude desejada.

Na Seção 2.6 uma outra técnica de transplante, chamada transplante por superposição, será mostrada. Essa técnica difere da anterior por manter certas ca-

racterísticas da sentença original modificando só parcialmente os valores de suas variáveis prosódicas.

2.2 Caracterização das Atitudes

Em seu artigo, MORAES [4] discorre sobre as características fonéticas e fonológicas de cada uma das 14 atitudes propostas abaixo. Posteriormente, LATSCH [1] utiliza as mesmas atitudes em sua tese e explica de maneira intuitiva as motivações e características de cada uma delas, e, seguindo a proposta de classificação de MORAES [4], as separa em três grupos: atitudes com características de subida de *pitch*, atitudes com características de descida de *pitch* e ainda atitudes que apresentem alterações marcantes na duração. Esses mesmos grupos serão mantidos neste trabalho para facilitar a classificação das atitudes.

Abaixo estão descritas as 14 diferentes atitudes trabalhadas por MORAES [4], além de uma descrição intuitiva adaptada de LATSCH [1] seguida de exemplos de aplicações em sentenças completas, com o objetivo de melhor ilustrar o uso dessas atitudes. Deve-se atentar para o fato de que as sentenças não se encontram em sua totalidade na norma culta do português brasileiro, uma vez que as sentenças em destaque dentro de cada contexto foram utilizadas na forma em que se apresentam e o contexto gerado em um momento posterior.

- Atitude Óbvia (*self-evident*): É utilizada em sentenças cujo orador considera que a informação é de conhecimento comum, como em: “É claro que **Renata jogava.**”;
- Sugestão (*suggestion*): É uma afirmativa na qual o orador indica uma ação a ser tomada de maneira não imperativa, como no caso: “Quer a minha opinião? **Repete de novo.**”;
- Ênfase Contrastiva (*contrastive emphasis*): É utilizada quando o orador deseja apresentar duas opções em uma sentença enquanto destaca uma das opções, como no caso a seguir: “Renata olhava ou **Renata jogava?**”;

- **Questão QU (*wh-question*):** Essa atitude se refere a sentenças interrogativas normalmente começadas por palavras iniciadas pelas letras QU, como por exemplo ‘Quem’, ‘Que’, ‘Qual’ e ‘Quando’. O termo QU foi adaptado de *wh-question* pois na língua inglesa essa atitude normalmente está associada a palavras como ‘Why’, ‘Who’, ‘Where’, ‘When’. Embora a atitude tenha sido nomeada fazendo referência às letras QU há ainda outras palavras que podem iniciar as sentenças dessa maneira como ‘Porque’, ‘Onde’ e ‘Como’. Um exemplo de questão QU seria: “Onde **Renata jogava?**”;
- **Comando (*command*):** É utilizada quando o orador deseja dar uma ordem a alguém utilizando uma entonação imperativa. Um exemplo de sentença com atitude Comando seria: “Volte lá agora e **Repete de novo!**”;
- **Exclamação QU (*wh-exclamation*):** Atitude que visa fazer uma afirmação de maneira impactante ou exagerada, mas que inicia por uma palavra QU, um exemplo seria: “Que grande bobagem!”;
- **Questão SN (*yes-no question*):** Essa atitude é utilizada quando o orador já sabe a resposta da questão que vai fazer e propõe a pergunta demonstrando isso. Essa questão não necessariamente requer uma resposta. Um exemplo se segue: “Preste atenção ao que vai responder à minha próxima pergunta. **Renata jogava?**”;
- **Questão SN Retórica (*rhetoric yes-no question*):** É a atitude utilizada quando o orador deseja provocar uma discussão fazendo uma pergunta, possivelmente sem resposta, sobre a qual já possui opinião formada. Um exemplo do uso dessa atitude seria: “Nesse contexto, podemos realmente dizer que **Renata jogava?**”;
- **Pedido (*request*):** Essa atitude é utilizada quando o orador deseja fazer uma solicitação não imperativa a alguém, como ilustrado a seguir: “O senhor poderia me informar se **Renata jogava?**”;
- **Questão SN Incrédula (*incredulous yn question*):** Essa atitude se apresenta quando o orador deseja realizar uma pergunta destacando uma certa característica, mas deixa claro o fato de que conhece a resposta previamente e esta

é o oposto daquilo que destacou na pergunta. Um exemplo de aplicação dessa atitude é: “Tem certeza de que é nisso que você quer que eu acredite, que **Renata jogava?**”;

- Aviso (*warning*): É uma atitude utilizada quando se deseja transmitir a necessidade de alerta, de atenção ou prevenção de algo. Uma possível aplicação: “Antes de ir para lá fique sabendo que era lá que **Renata jogava**”;
- Irônica (*ironic*): Utiliza-se essa atitude quando se deseja afirmar algo demonstrando que sua opinião é contrária aquilo que está sendo dito, ou que aquela afirmação é uma inverdade. Um caso de uso pode ser visto em: “**Renata jogava**, claro e nesse mesmo dia choviam porcos...”;
- Incrédula (*incredulous assertion*): Pode-se ouvir essa atitude quando o orador deseja apresentar uma afirmação demonstrando que não acredita no conteúdo daquilo que fala, como em: “Tem certeza de que **Renata jogava?**, pois não foi isso que me disseram”;
- Ênfase Intensiva (*intensive emphasis*): Essa atitude é utilizada quando o orador deseja mostrar que algo acontece de maneira mais intensa (forte) do que de costume ou que algo ocorre de maneira acima da média, como ocorre em: “**Renata jogava** e jogava e jogava sem nunca parar”;

2.3 Leitura dos Gráficos

A partir da próxima seção e ao longo de todo esse trabalho serão apresentados gráficos contendo a representação das características prosódicas de intensidade, *pitch* e duração das sentenças estudadas. Esses gráficos não possuem eixos ou unidades em sua apresentação neste trabalho, uma vez que são mostrados aqui não para se ter valores absolutos das características neles representados, mas tão somente uma comparação entre as amplitudes ao longo das sentenças.

Conforme pode ser observado na Figura 2.1, que contém gráficos para exemplificação, o gráfico mais à esquerda apresenta no topo a transcrição fonética da

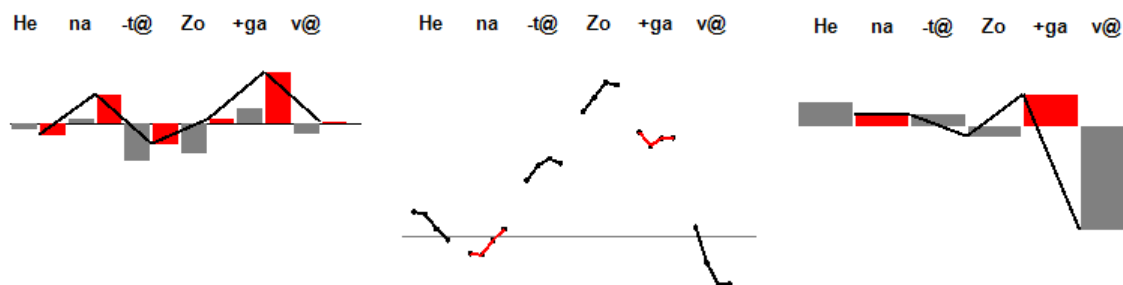


Figura 2.1: Exemplo de representação gráfica das características prosódicas das atitudes.

sentença representada, na parte inferior são apresentadas duas barras por sílaba da sentença, essas barras representam a variação da intensidade sonora referente a um valor padrão, para trechos que possuem intensidade superior ao valor padrão as barras estarão para cima da linha horizontal, sendo tão maiores quanto mais distantes desse valor sua intensidade for. Para trechos cuja duração é inferior a ao valor padrão as barras aparecerão para baixo da linha horizontal, sendo tão maiores quanto maior for a sua diferença de intensidade em relação a esse valor de referência da sentença.

Na mesma figura o gráfico central representa o *pitch* da sentença, no topo da figura aparece novamente a transcrição fonética da sentença. Na parte inferior são apresentados 4 pontos por sílaba, esses pontos representam os valores médios de *pitch* de cada trecho da sílaba, em relação a um valor padrão pré-estabelecido. Caso o ponto esteja numa posição superior à linha horizontal (referente ao valor padrão), isso significa um valor de *pitch* superior ao valor padrão, sendo tão mais alto quanto mais distante desse valor for o *pitch* médio do trecho. Para valores inferiores ao valor de referência os pontos estarão abaixo da linha, ficando tão mais para baixo quanto menores forem os valores de *pitch* médio nesse trecho.

No gráfico mais à direita estão representadas as durações de cada sílaba da sentença. É apresentada no topo da figura a transcrição fonética da sentença, na parte inferior do gráfico é apresentada uma barra de duração para cada sílaba. Caso a barra esteja apontando para cima, isso significa uma duração da sílaba maior que um valor de referência, sendo as barras tão maiores quanto maior for a duração da

sílaba. Caso a barra aponte para baixo isso significa uma duração menor que o valor de referência, sendo tão menores quanto maiores forem as barras apontando para baixo.

2.4 Caracterização Prosódica das Atitudes

A seguir será feita a análise das 14 diferentes estruturas quanto às suas características de pitch, duração e intensidade utilizando como base as análises feitas por MORAES [4] e apresentadas novamente em LATSCH [1].

2.4.1 Atitudes com queda de *pitch*

Nesse grupo estão inclusas seis das catorze atitudes classificadas por MORAES [4]. Entretanto haveria ainda uma sétima atitude nesse grupo, que foi aqui excluída pois LATSCH [1] a considerou de difícil reprodução, já que essa atitude teve uma baixa taxa de reconhecimento no trabalho original.

2.4.1.1 Óbvia

Essa atitude se diferencia da atitude de prosódia neutra por sua curva de *pitch* ascendente até a última sílaba tônica. Além disso, o alto valor do *pitch* na sua última sílaba tônica é o responsável pela sua identificação entre outras atitudes semelhantes. Outra característica que se observa é a elevação da intensidade sonora acima da média nas sílabas tônicas e postônicas finais, confirmando o fato de que essa é a região chave para a identificação da atitude, como pode ser visto na Figura 2.2 abaixo.

2.4.1.2 Sugestão

A atitude de sugestão possui características muito próximas às da atitude óbvia. Entretanto, para que a sentença tenha a característica de sugestão, seu conteúdo deve ser de um teor imperativo. Portanto, não há uma diferenciação clara entre essas duas atitudes no tocante às suas características prosódicas, conforme pode ser visto na Figura 2.3;

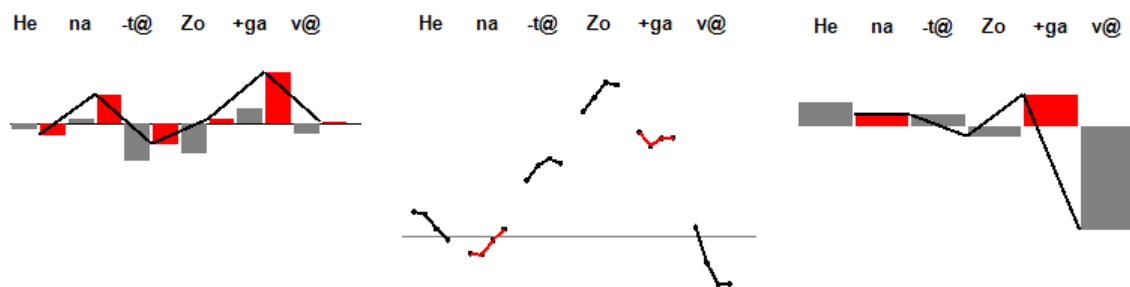


Figura 2.2: Da esquerda para a direita: Gráficos de Intensidade, *Pitch* e duração da sentença “Renata Jogava.” com atitude Óbvia.

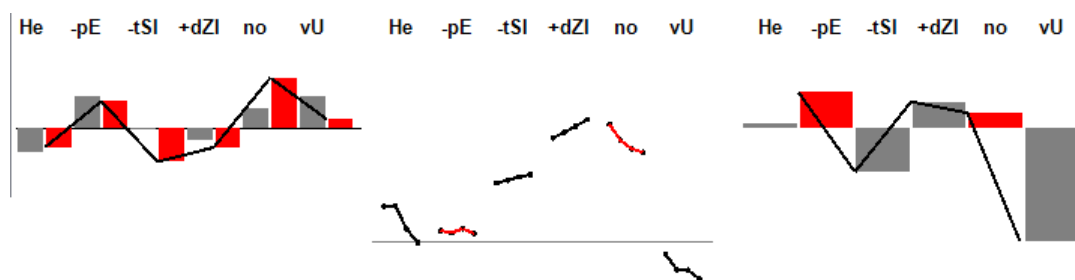


Figura 2.3: Da esquerda para a direita: Gráficos de Intensidade, *Pitch* e duração da sentença “Repete de novo.” com atitude Sugestão.

2.4.1.3 Ênfase Contrastiva

As características notórias dessa atitude, que podem ser observadas na Figura 2.4, frente à atitude neutra são o alto valor do *pitch* da sílaba pretônica final seguido de um formato convexo e de baixo valor na curva de *pitch* da sílaba tônica final. Segundo MORAES [4], apesar de essas características definirem completamente o *pitch*, elas não são suficientes para que a atitude tenha uma boa taxa de identificação, de fato somente quando são modificadas também a duração e a intensidade da sentença é que o reconhecimento chega a 80%.

2.4.1.4 Questão QU

Para analisar esta atitude foi necessário utilizar uma sentença diferente daquela utilizada anteriormente (“Renata jogava”). Substituindo a sentença por uma com uma palavra QU que mantivesse a estrutura silábica, como por exemplo “e como jo-

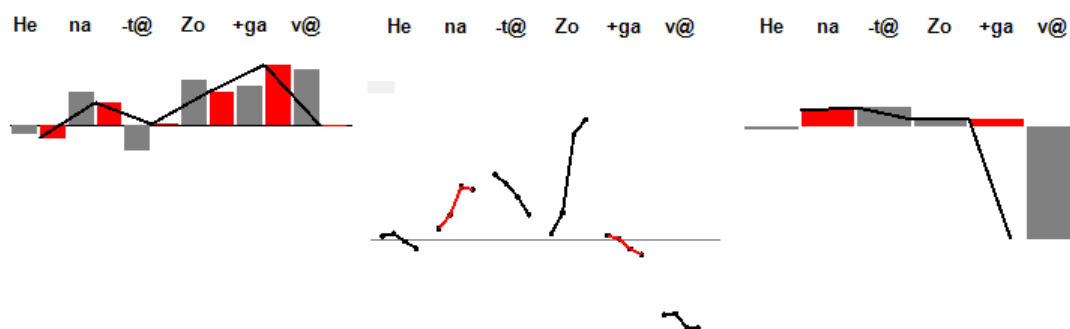


Figura 2.4: Da esquerda para a direita: Gráficos de Intensidade, *Pitch* e duração da sentença “Renata Jogava.” com atitude Ênfase Contrastiva.

gava” (Figura 2.5), observou-se um alto valor de *pitch* na palavra QU e uma posterior descida do *pitch* até a última sílaba tônica.

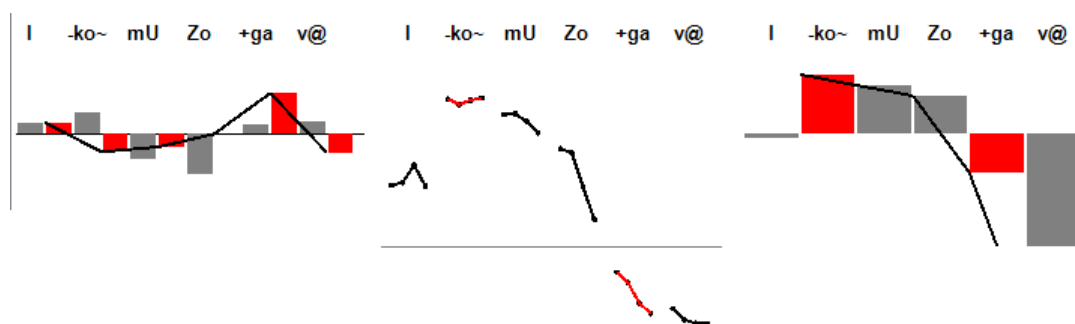


Figura 2.5: Da esquerda para a direita: Gráficos de Intensidade, *Pitch* e duração da sentença “E como jogava.” com atitude Questão QU.

2.4.1.5 Comando

As curvas de *pitch*, intensidade e duração dessa atitude, que se encontram na Figura 2.6, possuem características bastante similares às da atitude Questão QU, de forma que mais uma vez o que diferencia a interpretação dessa atitude frente à outra já citada é o conteúdo da sentença.

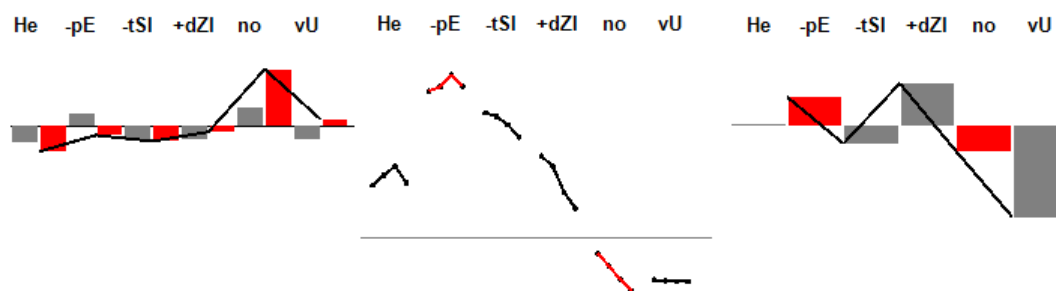


Figura 2.6: Da esquerda para a direita: Gráficos de Intensidade, *Pitch* e duração da sentença “Repete de novo.” com atitude Comando.

2.4.1.6 Exclamação QU

Essa atitude também possui características prosódicas semelhantes às aquelas encontradas na atitude Questão QU. Entretanto, diferentemente da atitude de comando, há uma diferenciação na sílaba pretônica final, onde há um valor de *pitch* mais baixo do que na Questão QU, que torna essa atitude facilmente perceptível, o que pode ser observado na Figura 2.7.

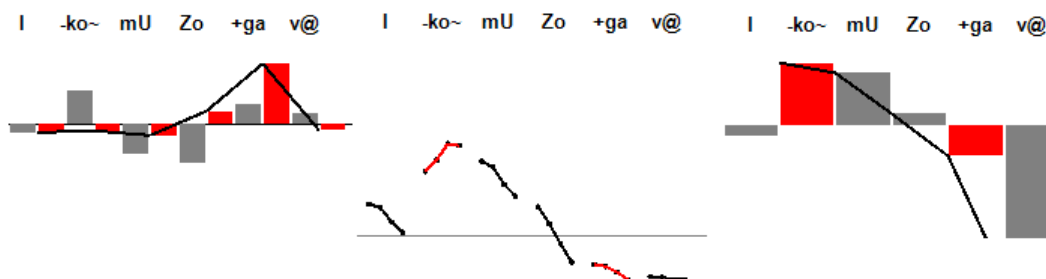


Figura 2.7: Da esquerda para a direita: Gráficos de Intensidade, *Pitch* e duração da sentença “E como jogava.” com atitude Exclamação QU.

2.4.2 Atitudes com subida de pitch

2.4.2.1 Questão SN

A curva de *pitch* dessa atitude (Figura 2.8) se caracteriza por uma subida na sílaba tônica inicial seguida de uma queda ao chegar na sílaba postônica inicial, em seguida, há uma subida na sílaba tônica final com uma posterior queda na sílaba

postônica final. Quanto aos valores de intensidade e duração, não há diferenças notáveis quando a comparamos à atitude neutra.

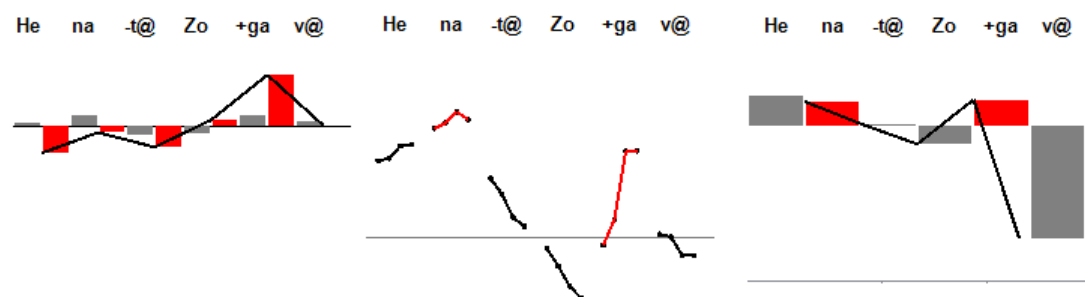


Figura 2.8: Da esquerda para a direita: Gráficos de Intensidade, *Pitch* e duração da sentença “Renata Jogava.” com atitude Questão SN.

2.4.2.2 Questão SN Retórica

Nessa atitude pode-se observar (Figura 2.9) uma curva de *pitch* muito semelhante àquela que existe na atitude Questão SN. Entretanto a Questão SN Retórica possui na sílaba tônica final uma queda, ao passo que na Questão SN essa queda se dá na sílaba postônica. Outra diferença notável é o menor valor de *pitch* da sílaba pretônica final nessa atitude.

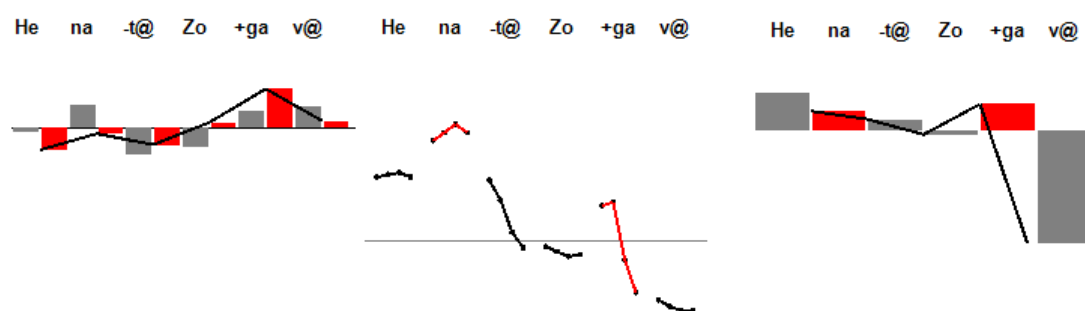


Figura 2.9: Da esquerda para a direita: Gráficos de Intensidade, *Pitch* e duração da sentença “Renata Jogava.” com atitude Questão SN Retórica.

2.4.2.3 Pedido

Essa atitude possui característica prosódicas (mostradas na Figura 2.10) muito semelhantes às da atitude Questão SN Retórica, não possuindo nenhuma característica marcante que faça a desambiguação entre essas duas atitudes. Para se identificar qual das atitudes está sendo usada, deve-se atentar ao conteúdo, que é na maior parte dos casos bastante diferente.

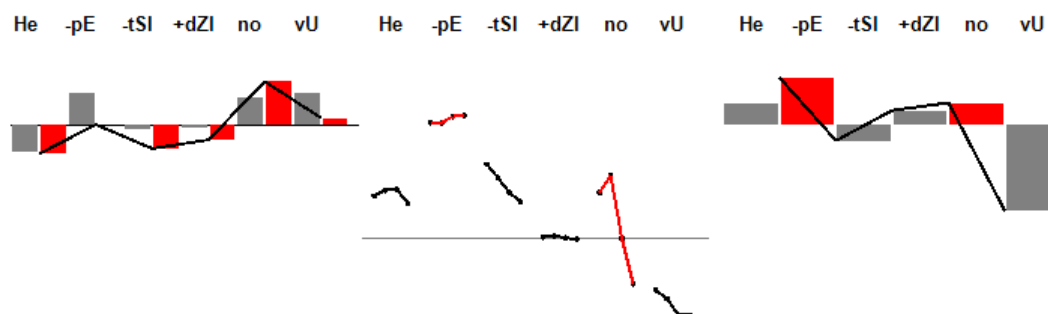


Figura 2.10: Da esquerda para a direita: Gráficos de Intensidade, *Pitch* e duração da sentença “Repete denovo.” com atitude Pedido.

2.4.2.4 Questão SN Incrédula

As características prosódicas dessa atitude se mostram na forma de uma curva de *pitch* com um vale formado pela descida do *pitch* na sílaba pretônica seguida de uma subida na sílaba tônica final. Afora essas características citadas, não há nenhuma outra característica de intensidade ou duração que seja notável, o que se observa na Figura 2.11.

2.4.3 Atitudes com variação de duração

2.4.3.1 Aviso

A atitude de aviso tem sua caracterização no alongamento da vogal da sílaba tônica final realizado em conjunto com um aumento na intensidade sonora dessa vogal. Além disso, há um valor bastante alto do *pitch* na sílaba pretônica final. As curvas que caracterizam essa atitude podem ser vistas na Figura 2.12.

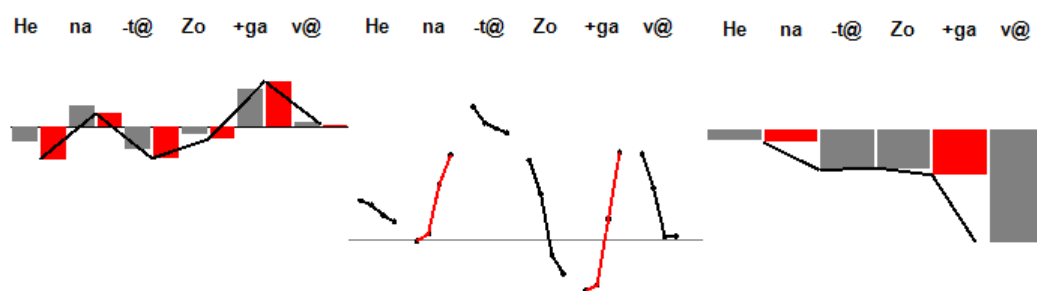


Figura 2.11: Da esquerda para a direita: Gráficos de Intensidade, *Pitch* e duração da sentença “Renata Jogava.” com atitude Questão SN Incrédula.

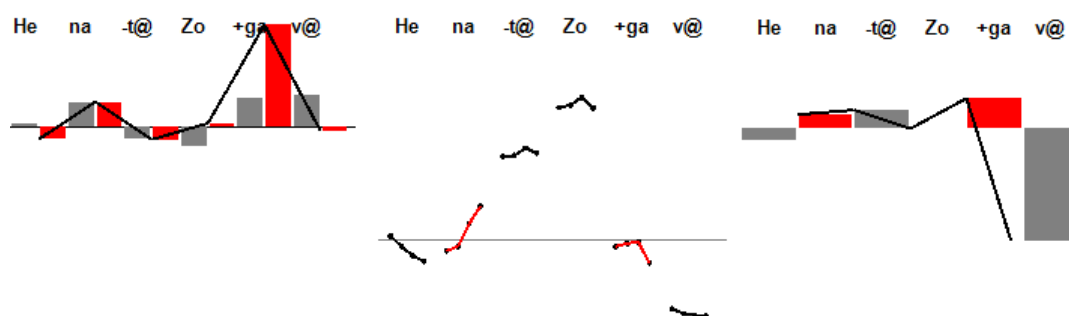


Figura 2.12: Da esquerda para a direita: Gráficos de Intensidade, *Pitch* e duração da sentença “Renata Jogava.” com atitude Aviso.

2.4.3.2 Irônica

Essa atitude se caracteriza por uma curva de *pitch* e duração muito parecida com a da atitude Aviso (Figura 2.13), com aumento da duração da sílaba tônica final e alto valor de *pitch* na sílaba pretônica. Entretanto, como diferenciação entre as duas atitudes, pode-se notar que na atitude irônica há a manutenção de um valor constante até o meio da vogal da tônica final, com um valor bastante inferior ao da sílaba anterior, seguido de um crescimento do *pitch*, o que torna sua assinatura bastante característica.

2.4.3.3 Incrédula

As curvas de *pitch* (Figura 2.14) dessa atitude são pouco variadas, apresentando quase nenhuma subida ou declínio ao longo da sentença. Dessa forma, essa característica de *pitch flat* se torna um fator de forte identificação dessa atitude. No

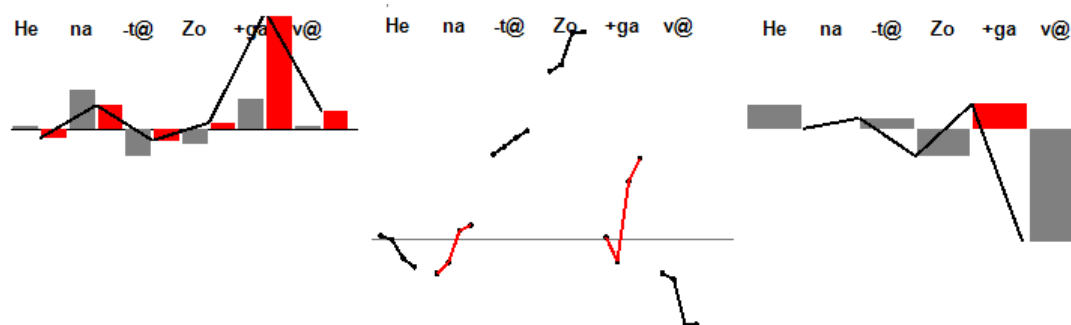


Figura 2.13: Da esquerda para a direita: Gráficos de Intensidade, *Pitch* e duração da sentença “Renata Jogava.” com atitude Irônica.

escopo da duração pode-se notar um aumento em todas as consoantes, notadamente na consoante da sílaba tônica final.

LATSCH [1] aponta o fato de que há poucas diferenças entre os parâmetros prosódicos dessa atitude e da atitude Neutra, de forma que a ressíntese das sentenças no método de transplante por superposição acaba por descaracterizar a atitude. Nesse caso, melhores resultados são obtidos com o uso da técnica de transplante por cópia.

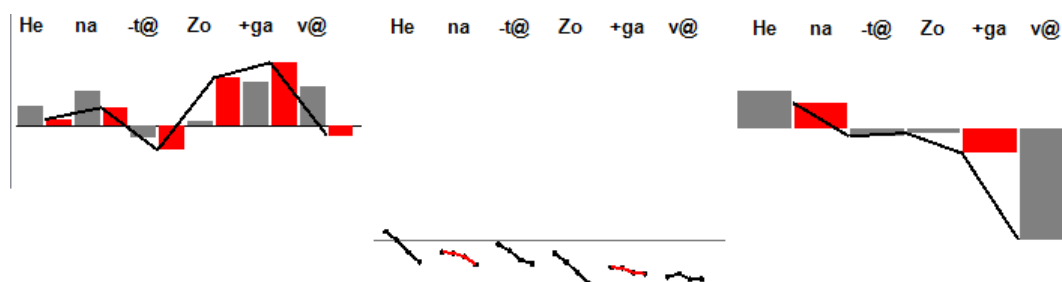


Figura 2.14: Da esquerda para a direita: Gráficos de Intensidade, *Pitch* e duração da sentença “Renata Jogava.” com atitude Incrédula.

2.4.3.4 Ênfase Intensiva

Os fatores determinantes para se identificar sentenças com essa atitude são o alto nível do *pitch* no fim da sílaba tônica, com a manutenção desse nível até o fim da

sílaba postônica e um aumento da duração das vogais nessas sílabas, como visto abaixo na Figura 2.15.

LATSCH [1] também observa que esse tipo de atitude caracteriza uma sentença não terminada, o que explica o fato de haver valores altos de *pitch* no final da sentença.

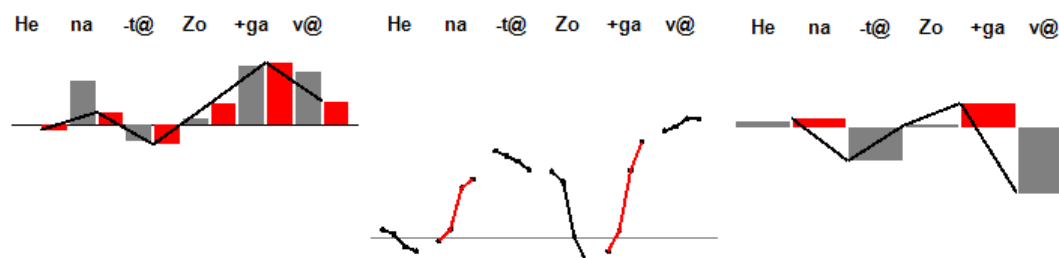


Figura 2.15: Da esquerda para a direita: Gráficos de Intensidade, *Pitch* e duração da sentença “Renata Jogava.” com atitude Ênfase Intensiva.

Uma característica de todas as atitudes ressaltada por LATSCH é o fato de que em poucas delas a intensidade tem um papel muito relevante na identificação da atitude. Uma possível causa disso é o fato de que a intensidade pode variar muito entre oradores e entre diferentes gravações. Além disso, na maioria dos casos, podemos notar que o aumento da intensidade está associado ao aumento da duração das sílabas e vogais.

2.5 Transplante de prosódia por cópia

Considerando a normalização das variáveis prosódicas dentro de uma sílaba, LATSCH [1] apresenta uma forma de se fazer o transplante de prosódia entre duas sentenças cujas sílabas possuam conteúdos distintos. Esta técnica é chamada de “transplante de prosódia por cópia”. Embora o conteúdo das sílabas possa ser diferente, essa técnica requer uma estrutura silábica idêntica nas duas sentenças, já que os parâmetros atribuídos a cada sílaba são fortemente dependentes de sua posição na sentença.

Neste contexto, a expressão “estrutura silábica” refere-se ao número de sílabas numa sentença bem como a posição das sílabas tônicas na mesma. De maneira ilustrativa, a sentença “Renata Jogava”, que será doravante muito utilizada nesse trabalho, possui a mesma estrutura silábica da sentença “Brinquedo Quebrava”, pois ambas apresentam seis sílabas e ambas as palavras de cada uma das sentenças são paroxítonas.

A técnica do transplante de prosódia por cópia é uma das mais simples dentre as técnicas de transplante encontradas na literatura (como em LATSCH [1]): ela se dá a partir da cópia dos parâmetros prosódicos de cada uma das sílabas da sentença de teste para a sua sílaba equivalente na sentença de referência, de modo que as curvas de *pitch*, intensidade sonora e duração de cada uma das sílabas se tornem iguais nas duas sentenças.

Conforme ilustrado na Figura 2.16, o processo do transplante por cópia se dá da seguinte maneira: Inicialmente é feita a análise de ambas as sentenças, com a extração das características de cada uma, como as etiquetas e as marcas de *pitch*. A seguir, esses parâmetros são passados para um método de análise silábica que relaciona os dados obtidos na etapa anterior com as marcações das sílabas em cada palavra. Na etapa seguinte, os parâmetros da sentença de teste são copiados para a sentença de referência. Por fim, há a síntese das sílabas com os novos parâmetros copiados da sentença de teste, através de algoritmo do tipo PSOLA (Pitch Synchronous Overlap and Add).

Um exemplo prático do transplante por cópia é mostrado abaixo na Figura 2.17 onde se podem ver as curvas de pitch da sentença de teste “Renata Jogava” dita com a atitude de Aviso (*warning*, explicada na Seção 2.4.3.1) e da sentença “Cristina Lembrava”, que neste caso atua como a sentença de referência.

Pode-se notar facilmente que não há correspondência temporal direta entre os trechos de cada uma das sentenças usadas. Há, porém, a manutenção da estrutura silábica, de forma que, ao final do procedimento de transplante, as curvas de *pitch*

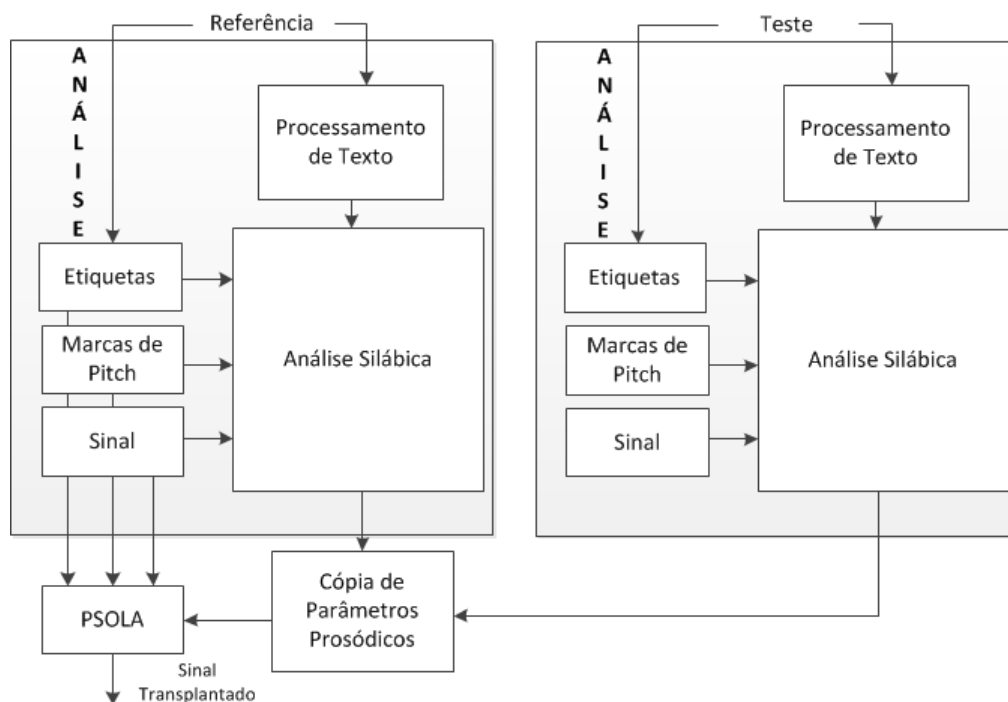


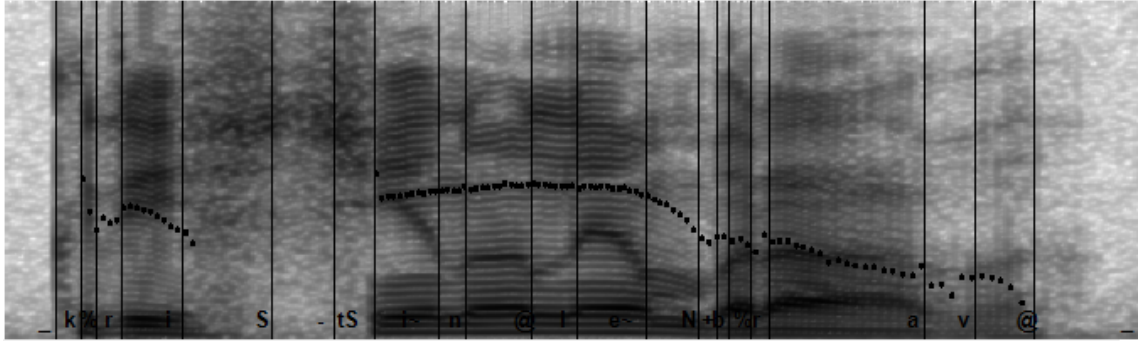
Figura 2.16: Esquemático do transplante por cópia.

da sentença de teste e da sentença modificada são aproximadamente iguais, a menos de modificações oriundas da duração e conteúdo da sentença de referência.

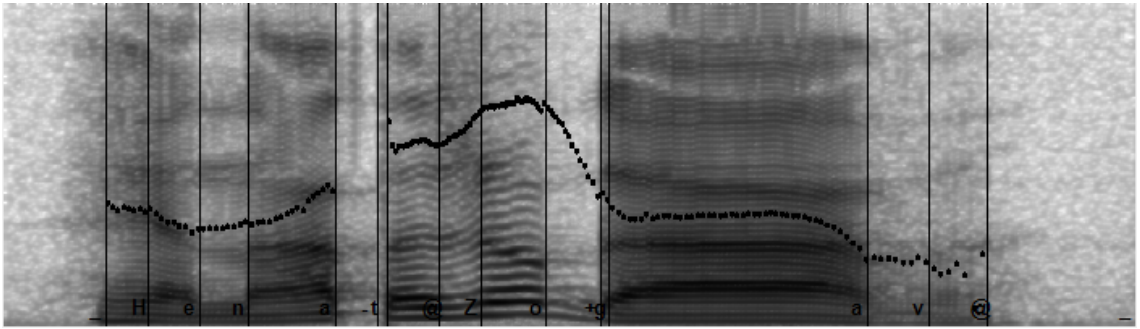
2.6 Transplante de prosódia por superposição

Alternativamente à técnica de transplante por cópia, LATSCH [1] propõe também uma técnica de transplante que não mais copia indistintamente os parâmetros prosódicos de cada sílaba da sentença de teste para a sentença de referência. De fato, esta outra técnica assume haver curvas de *pitch* e duração básicas para uma dada sentença, como pode ser observado no modelo SFC [5], e que a atitude é um padrão que deve ser adicionado às curvas básicas para se obter a sentença dita com a entonação desejada. Dessa forma, há a manutenção das características próprias de cada sentença, bem como das características vindas da atitude desejada.

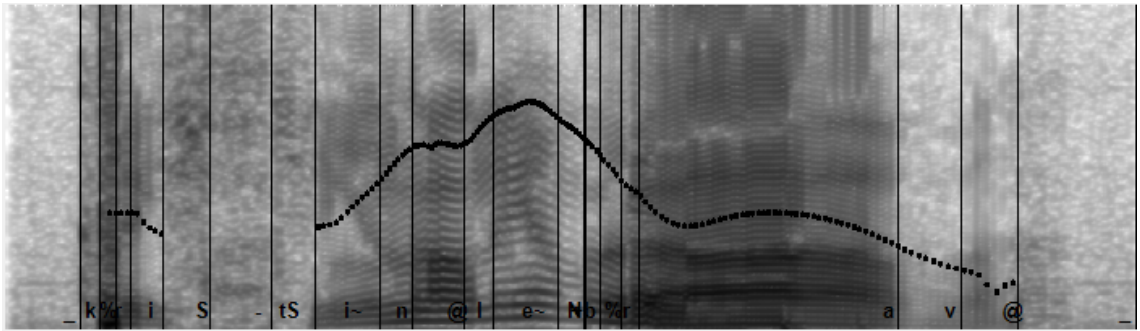
Para que seja utilizada essa técnica deve-se notar que ambas as sentenças trabalhadas devem possuir a mesma estrutura silábica. Isso porque tal procedimento pressupõe uma relação direta entre a posição da sílaba na frase e suas características



(a) Sentença de Referência: “Cristina lembrava” com atitude Neutra.



(b) Sentença de teste: “Renata jogava” com atitude Aviso.



(c) Resultado do transplante.

Figura 2.17: Exemplo do transplante por cópia.

de *pitch* e duração, sendo então essas características transportadas da sentença de teste para a de referência.

Com o intuito de se efetuar o transplante por superposição, deve-se obter inicialmente as gravações de três sentenças: a sentença de referência, a qual desejamos modificar, a sentença de teste, dita em prosódia neutra, e a sentença de teste com a prosódia desejada, por exemplo dita de maneira óbvia. Em seguida, obtêm-se os parâmetros prosódicos de cada sílaba de cada sentença, conforme o método utilizado no transplante por cópia. Uma vez obtidos os parâmetros, subtraem-se os valores

dos parâmetros prosódicos da sentença de teste dita em prosódia neutra da sua equivalente na prosódia que se deseja simular, para obter a modificação desejada. A seguir somam-se, sílaba a sílaba, os valores dos parâmetros da sentença de referência com aqueles obtidos pela subtração feita anteriormente. Esses parâmetros são então passados a um outro módulo, onde é feita a síntese das sílabas e posterior utilização do PSOLA para fazer as modificações necessárias no sinal. Dessa forma, podemos dizer que não houve uma cópia dos parâmetros prosódicos nas sentenças, mas uma transformação da sentença de referência. Como pode ser observado na Figura 2.18.

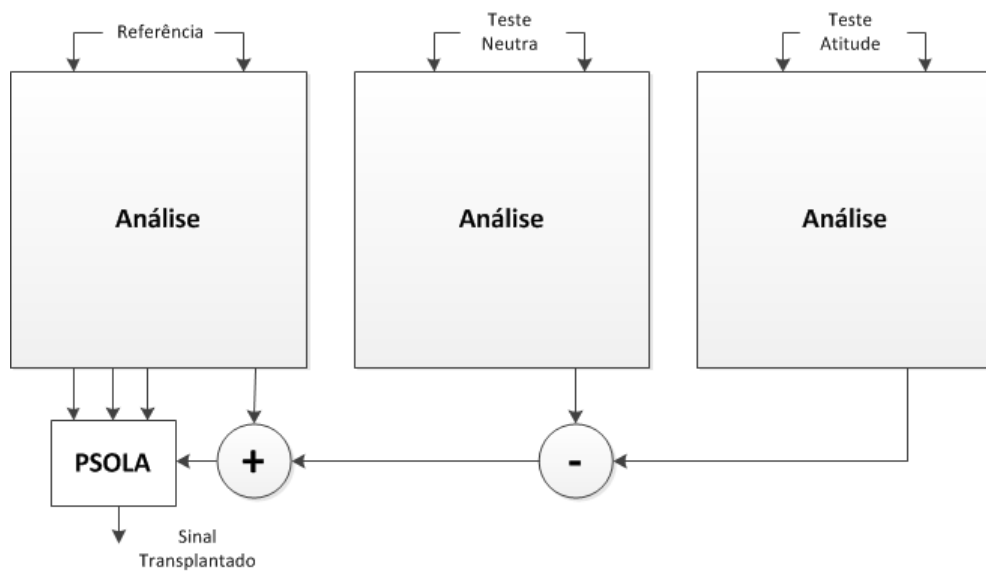


Figura 2.18: Esquemático do transplante por superposição.

2.7 Conclusões

Ao longo do capítulo foram apresentados dois diferentes métodos de transplante de prosódia, além de uma análise das propriedades que caracterizam cada atitude individualmente.

No primeiro método, chamado de transplante por cópia, os valores das variáveis de prosódia de uma sentença foram forçosamente copiados para a outra sentença, fazendo assim com que a segunda apresentasse exatamente as mesmas durações, intensidades e *pitch* da primeira. Esse método se mostrou eficaz para a maior parte

das atitudes estudadas.

A seguir, foi apresentado um outro método chamado transplante por superposição, que extrai as variáveis prosódicas de uma dada atitude obtendo as diferenças entre uma mesma sentença dita com a atitude e de maneira neutra. Posteriormente, este outro método adiciona essas variáveis extraídas às variáveis que caracterizam a atitude neutra de outra sentença. Essa técnica se mostrou eficaz na maior parte dos casos. Entretanto houve casos, como por exemplo, a atitude incrédula, em que o transplante por cópia apresentava resultados melhores, por captar certas nuances que a superposição executava com imperfeição.

Embora os dois métodos apresentados funcionem bastante bem para sentenças com a mesma estrutura silábica, por suas concepções eles não são capazes de trabalhar com sentenças de diferentes estruturas, de modo que isso limita bastante o seu uso. Um dos problemas dessa limitação é o fato de se precisar manter um banco de sentenças muito maior para poder contemplar a maior parte das sentenças com que se deseja trabalhar.

Quanto às análises das características fonéticas das atitudes que foram feitas no capítulo, pode-se notar que só foi utilizada uma estrutura silábica. Assim, é possível que haja algumas dessas características observadas que dependem somente da atitude, independente da estrutura a ser analisada.

No capítulo seguinte serão feitas novas análises das características de sentenças com estruturas diferentes daquelas apresentadas aqui, de forma a se obter mais informações acerca daquilo que realmente caracteriza uma atitude. Além disso, será feita uma discussão sobre um algoritmo capaz de transplantar uma atitude entre sentenças com estruturas silábicas distintas.

Capítulo 3

Classificação de Atitudes em Estruturas Silábicas Diferentes

3.1 Introdução

No capítulo anterior foram vistas técnicas de transplante de prosódia em sentenças com uma dada estrutura silábica. Essas técnicas, entretanto, devem ser ligeiramente adaptadas para uso em sentenças com outras estruturas silábicas.

Também no capítulo anterior foram analisados diversos sinais, com diferentes atitudes prosódicas e uma mesma estrutura silábica, e a partir disso um modelo para cada atitude foi gerado. Nesse capítulo o foco será a análise das atitudes em sentenças com diferentes estruturas silábicas daquela estudada no Capítulo 3.

Nesse sentido, na Seção 3.2 são analisados sinais com 14 diferentes contextos prosódicos e estruturas silábicas distintas daquela utilizada no capítulo anterior. Para esses sinais, são observadas as características de *pitch*, duração e intensidade sonora, de forma a observar como se comportam no tempo.

Na Seção 3.3 as características observadas na Seção 3.2 são confrontadas com o modelo obtido no Capítulo 2, para que esse modelo seja generalizado tornando-se válido para outras estruturas silábicas.

3.2 Caracterização Prosódica das Atitudes em Estrutura Silábica Alternativa

Nessa seção analisamos sentenças com as mesmas 14 atitudes discutidas na capítulo anterior, porém com uma estrutura silábica (em termos do posicionamento das sílabas tônicas) diferente. No caso, foi usada a sentença “Pífanos tocarão”, formada por uma palavra proparoxítona seguida de uma oxítona, ambas com 3 sílabas.

Para as atitudes do tipo QU usamos a sentença “Quando que passará” e para o teor imperativo usamos “Cala-te jogador”. Vale ressaltar nesse ponto que nem todas as sentenças estão de acordo com a norma culta do português brasileiro, mas serão aqui utilizadas por sua estrutura silábica mais do que pelo seu sentido semântico.

De forma a facilitar o entendimento e comparação com os casos estudados anteriormente, as atitudes prosódicas serão novamente divididas em 3 grupos (vistos separadamente nas subseções a seguir): atitudes com queda de *pitch*, atitudes com subida de *pitch* e atitudes com variação de duração.

3.2.1 Sinais com queda de *pitch*

3.2.1.1 Óbvvia

Na caracterização dessa atitude, ilustrada na Figura 3.1, podemos observar que o valor do *pitch* é crescente do início da sentença até a sílaba tônica final. Além disso, pode ser observado um valor mais alto da intensidade sonora nas sílabas tônicas, principalmente na sílaba tônica final, o que demonstra uma importância maior dessa sílaba para o reconhecimento dessa atitude.

3.2.1.2 Sugestão

A curva de *pitch* dessa atitude (Figura 3.2) tem suas características mais marcantes no crescimento do valor do *pitch* até a sílaba pretônica final. A partir da sílaba tônica final observa-se uma queda do *pitch*. Suas curvas de intensidade sonora e duração das sílabas não apresentam características marcantes.

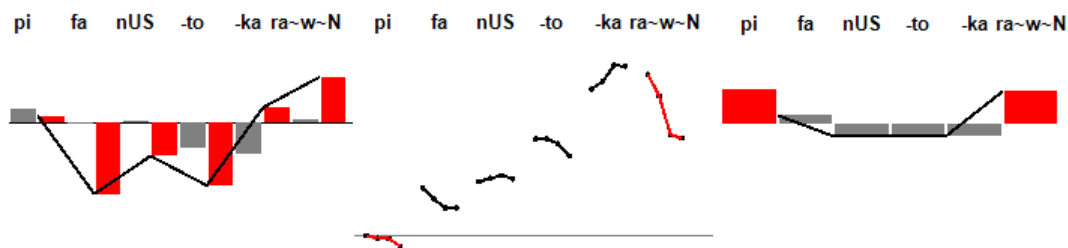


Figura 3.1: Da esquerda para a direita: Gráficos de Intensidade, *Pitch* e duração da sentença “Renata Jogava.” com atitude Óbvia.

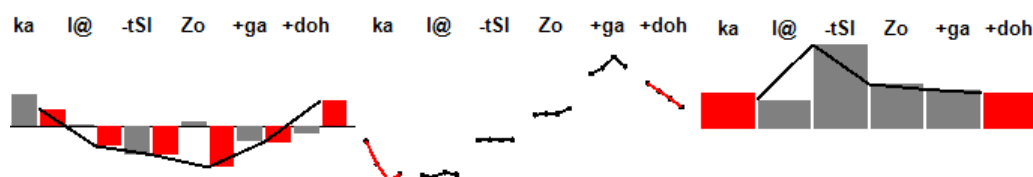


Figura 3.2: Da esquerda para a direita: Gráficos de Intensidade, *Pitch* e duração da sentença “Cala-te Jogador.” com atitude Sugestão.

3.2.1.3 Ênfase Contrastiva

A atitude de Ênfase Contrastiva caracteriza-se por um crescimento no *pitch* da sentença até a sílaba pretônica final, de modo que esta sílaba tem um *pitch* maior que o das outras sílabas da sentença. Além disso, de modo a destacar essa característica no *pitch* da sílaba pretônica, observa-se que sua duração é maior que a média das outras sílabas e a intensidade sonora nessa sílaba é maior quando comparada à atitude neutra, como pode ser observado abaixo na Figura 3.3.

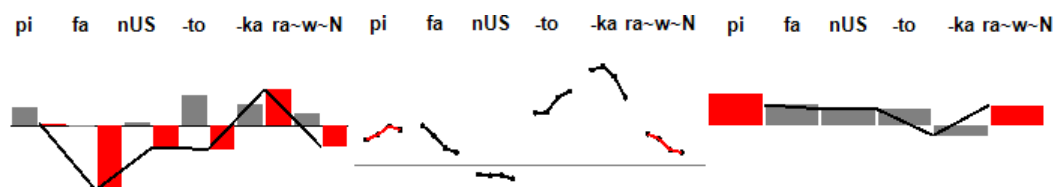


Figura 3.3: Da esquerda para a direita: Gráficos de Intensidade, *Pitch* e duração da sentença “Pífanos Tocarão.” com atitude Ênfase Contrastiva.

3.2.1.4 Questão QU

As principais características que facilitam a identificação da atitude nessa sentença são o alto valor do *pitch* na palavra QU, que nesse caso é “quando”, e um decrescimento desse valor até o fim da sentença. As curvas que caracterizam essa sentença podem ser observadas na Figura 3.4 (abaixo).

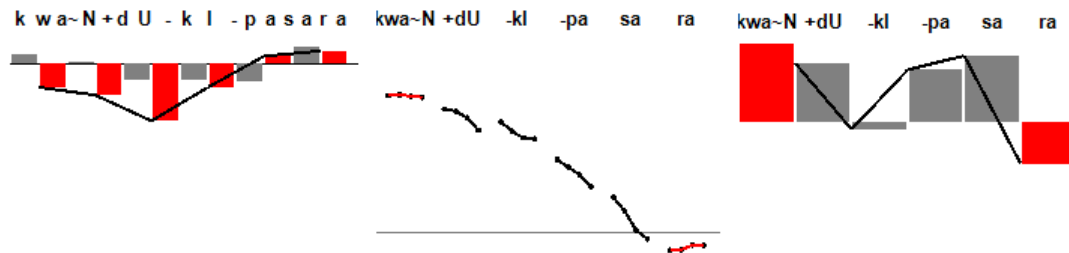


Figura 3.4: Da esquerda para a direita: Gráficos de Intensidade, *Pitch* e duração da sentença “Quando que Passará.” com atitude Questão QU.

3.2.1.5 Comando

A curva de *pitch* dessa sentença começa com uma queda em seu valor, a partir da tônica inicial, que segue até a sílaba pretônica final. Nesse ponto começa a subir, aumentando seu valor até o meio da sílaba tônica final. Observa-se, então, uma nova queda, como mostrado na figura 3.5.

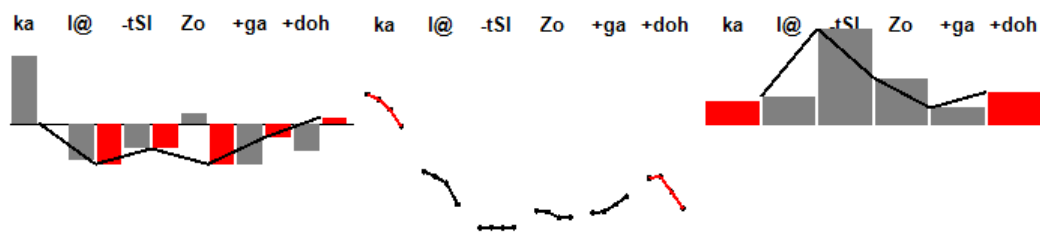


Figura 3.5: Da esquerda para a direita: Gráficos de Intensidade, *Pitch* e duração da sentença “Cala-te Jogador.” com atitude Comando.

3.2.1.6 Exclamação QU

As curvas características dessa sentença assemelham-se bastante às da sentença Questão QU, conforme pode ser observado na Figura 3.6, de modo que se observa um valor mais alto do *pitch* na palavra QU, seguido de um decrescimento nessa curva até o fim da sentença. O fator de desambiguação entre as duas atitudes está presente na sílaba pretônica final, onde se observa um valor de *pitch* mais baixo que na sentença com atitude Questão QU.

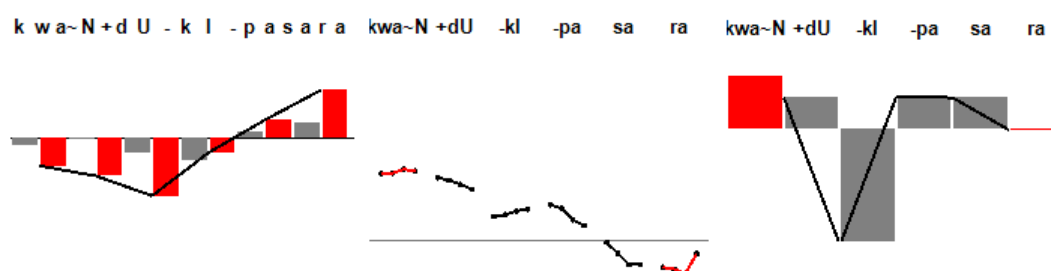


Figura 3.6: Da esquerda para a direita: Gráficos de Intensidade, *Pitch* e duração da sentença “Quando que Passará.” com atitude Exclamação QU.

3.2.2 Sinais com subida de pitch

3.2.2.1 Questão SN

Os valores de *pitch* nessa atitude destacam-se por serem crescentes na sílaba tônica inicial e em seguida decrescerem até a sílaba pretônica final. Na sílaba tônica final podemos observar novo crescimento dos valores de *pitch*, o que pode ser observado na Figura 3.7, que se refere a essa sentença. No escopo de intensidade, observa-se um valor acima da média nas sílabas tônica e pretônica finais.

3.2.2.2 Questão SN Retórica

As características de *pitch* dessa sentença aproximam-se bastante das observadas na sentença anterior, Questão SN. Entretanto, em vez de uma curva decrescente até a sílaba pretônica final e subida somente na tônica final, observa-se a subida já na sílaba pretônica final, fazendo assim com que haja uma diferença perceptível entre

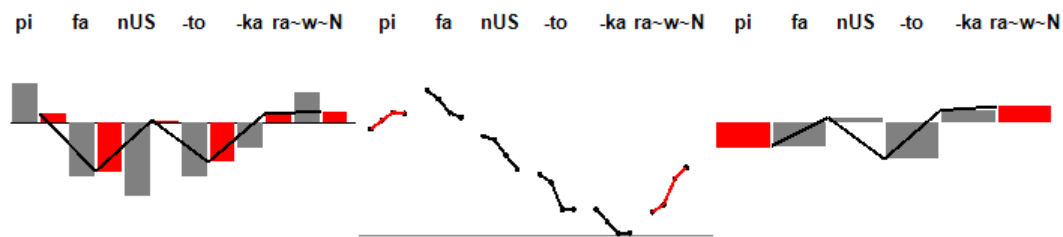


Figura 3.7: Da esquerda para a direita: Gráficos de Intensidade, *Pitch* e duração da sentença “Pifanos Tocarão.” com atitude Questão SN.

as duas atitudes. No contexto de intensidade e duração, as curvas novamente se assemelham às da atitude Questão SN, como se observa na Figura 3.8.

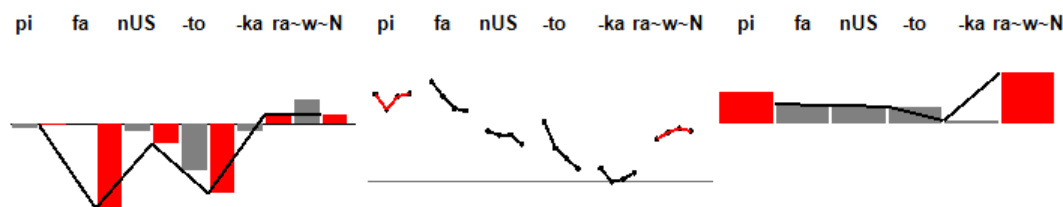


Figura 3.8: Da esquerda para a direita: Gráficos de Intensidade, *Pitch* e duração da sentença “Pifanos Tocarão.” com atitude Questão SN Retórica.

3.2.2.3 Pedido

Observa-se nessa sentença um valor de *pitch* subindo até a metade da sílaba tônica inicial, seguido de uma queda desse valor até o meio da sentença. A seguir, há um novo crescimento até a sílaba tônica final, quando, no fim da sílaba, observa-se nova queda no valor do *pitch*, o que dá às sílabas tônicas um formato convexo. As características dessa sentença podem ser observadas na Figura 3.9.

3.2.2.4 Questão SN Incrédula

Essa sentença tem sua atitude caracterizada por uma leve subida do *pitch* na sílaba tônica inicial, seguida de uma descida na sílaba pretônica final, que em conjunto com a subida na sílaba tônica final forma um vale na curva. Além dessas características,

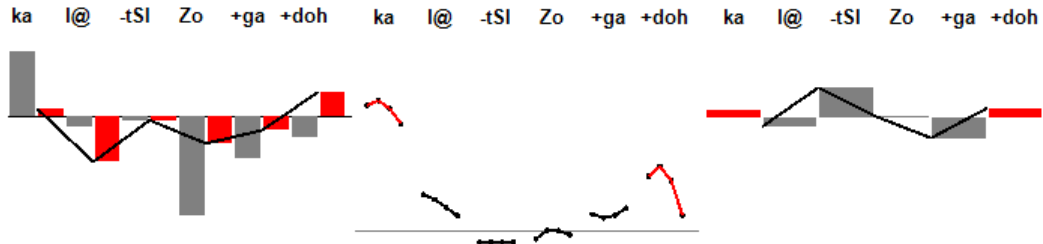


Figura 3.9: Da esquerda para a direita: Gráficos de Intensidade, *Pitch* e duração da sentença “Cala-te Jogador.” com atitude Pedido.

que podem ser observadas na Figura 3.10, pode-se notar um valor mais alto da intensidade sonora nas sílabas tônicas.

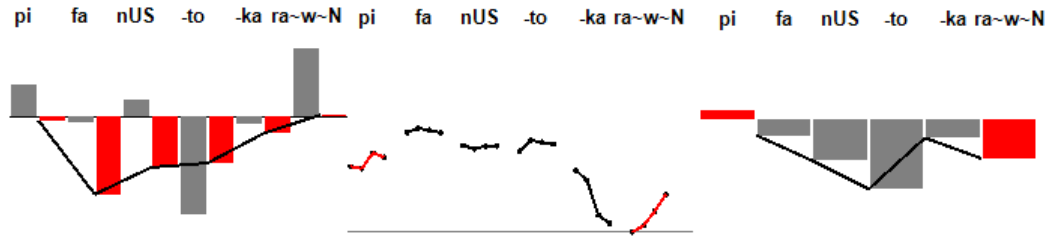


Figura 3.10: Da esquerda para a direita: Gráficos de Intensidade, *Pitch* e duração da sentença “Pífanos Tocarão.” com atitude Questão SN Incrédula.

3.2.3 Sinais com variação de duração

3.2.3.1 Aviso

Nessa sentença, observa-se um crescimento do valor de *pitch* até a sílaba pretônica final. No escopo de duração, observa-se um valor maior na sílaba tônica final. Além disso, no que diz respeito à intensidade sonora, também podemos observar um maior valor na sílaba tônica final, indicando que essa é a região chave para a caracterização da atitude nessa sentença. As curvas relacionadas a essa sentença encontram-se ilustradas na Figura 3.11.

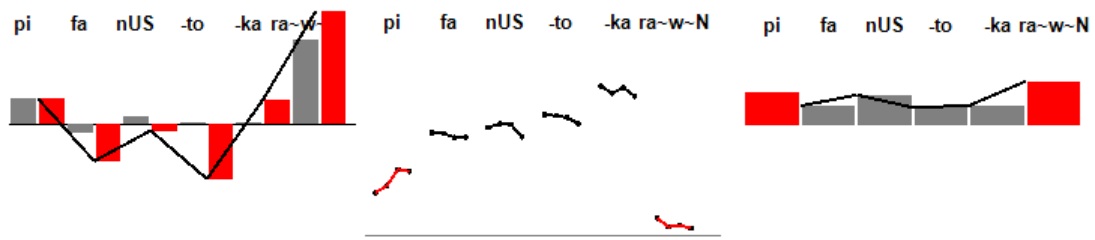


Figura 3.11: Da esquerda para a direita: Gráficos de Intensidade, *Pitch* e duração da sentença “Pífanos Tocarão.” com atitude Aviso.

3.2.3.2 Irônica

Assim como na sentença com atitude de aviso, essa sentença tem a curva de *pitch* crescente até a sílaba pretônica final (Figura 3.12). Entretanto, na duração o alongamento das sílabas acontece desde a sílaba pretônica final, e não somente na tônica.

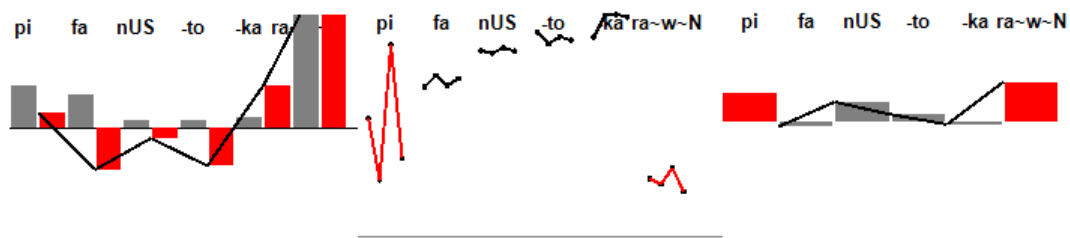


Figura 3.12: Da esquerda para a direita: Gráficos de Intensidade, *Pitch* e duração da sentença “Pífanos Tocarão.” com atitude Irônica.

3.2.3.3 Incrédula

A curva de *pitch* dessa sentença possui trechos praticamente constantes (*flat*), sofrendo uma leve queda ao longo do tempo. Quanto à duração, pode-se observar o aumento da curva nas sílabas tônica e pretônica final, conforme se vê na Figura 3.13.

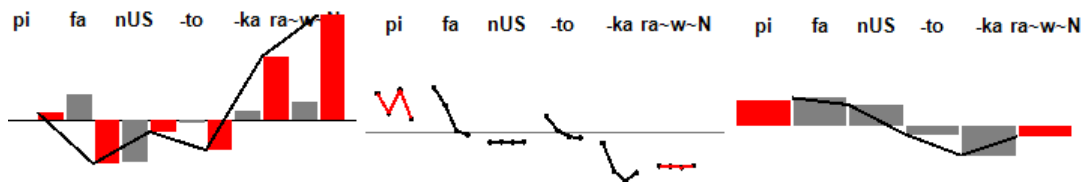


Figura 3.13: Da esquerda para a direita: Gráficos de Intensidade, *Pitch* e duração da sentença “Pifanos Tocarão.” com atitude Incrédula.

3.2.3.4 Ênfase Intensiva

As curvas que caracterizam essa sentenças assemelham-se bastante às da sentença com atitude incrédula. O *pitch* dessa sentença é constante em quase toda a duração do sinal, sofrendo uma queda na sílaba pretônica final e voltando a subir na tônica final. As curvas de duração e intensidade se mantêm iguais às da sentença com atitude incrédula, fato que pode ser visto na Figura 3.14.

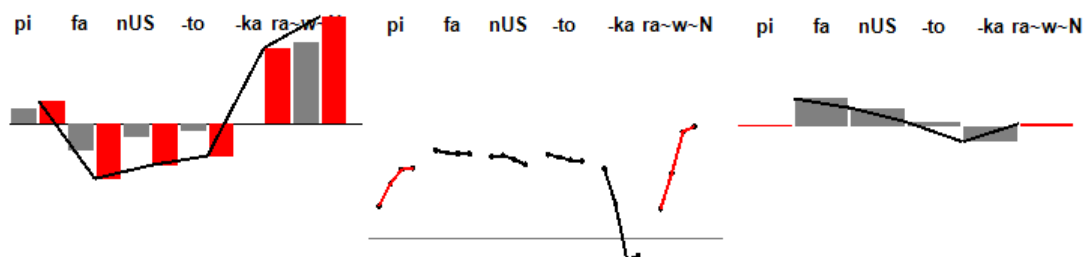


Figura 3.14: Da esquerda para a direita: Gráficos de Intensidade, *Pitch* e duração da sentença “Pifanos Tocarão.” com atitude Ênfase Intensiva.

3.3 Análise da Validade do Modelo

Na presente seção, comparamos os modelos das atitudes obtidos por MO-RAES [4] com as características apresentadas por estas mesmas atitudes para a nova estrutura silábica. Para melhor organizar os resultados de nossas análises, mais uma vez será mantida a segmentação das atitudes.

3.3.1 Atitudes com queda de *pitch*

3.3.1.1 Óbvvia

Para essa atitude o modelo de MORAES [4] prevê um crescimento do *pitch* do início até a sílaba tônica final, com alto valor de intensidade sonora na tônica final. Como observado na sessão passada, a curva de *pitch* dessa sentença é ascendente do início ao fim do sinal. Como a estrutura silábica dessa sentença caracteriza-se por possuir a primeira e a última sílabas tônicas, pode-se afirmar que essa curva comporta-se conforme o esperado pelo modelo. A outra característica relevante nesse sinal é a intensidade sonora se elevar nas sílabas tônicas. Observa-se no modelo um alto valor dessa variável na sílaba tônica final. Portanto, o modelo anterior pode ser considerado válido, para estruturas silábicas diferentes daquelas daquela através da qual foi feito.

3.3.1.2 Sugestão

O modelo original dessa atitude estabelece que as características de *pitch*, intensidade sonora e duração dessa atitude devem ser as mesmas da sentença Óbvvia, embora o conteúdo da sentença deva ser imperativo. Observa-se no sinal analisado que a curva de *pitch* segue crescente até a sílaba pretônica final, e na sílaba tônica final inicia uma queda, de modo que essa curva se encaixa nos moldes previstos para a atitude Óbvvia e, por conseguinte, no modelo da atitude Sugestão. Embora não se possa observar um valor maior na intensidade sonora desse sinal, esse fator não é determinante, pois há muita variação entre gravações nos valores da intensidade. Portanto pode-se afirmar que o modelo para essa atitude é válido, para outras estruturas diferentes daquela para o qual foi feito.

3.3.1.3 Ênfase Contrastiva

Para ser considerado válido o modelo dessa atitude, a sentença deve apresentar um alto valor de intensidade sonora na sílaba tônica final. Sua curva de *pitch* deve ser crescente até a sílaba pretônica final, que tem um valor alto. A sílaba que se segue, ou seja a tônica final, apresenta um valor mais baixo e deve possuir uma forma convexa. No sinal analisado, vê-se um comportamento adequado a essas

características e, embora o valor de da intensidade sonora na tônica final não seja o maior da sentença, pode-se observar que o mesmo é maior que o observado na sentença dita com atitude neutra. Considerando todos esses fatores, é adequado assumir que o modelo para essa atitude é válido, apesar da estrutura silábica.

3.3.1.4 Questão QU

O modelo dessa atitude é bastante simples, necessitando somente de duas propriedades para caracterizá-lo: a primeira é um alto valor de *pitch* na palavra QU e a segunda uma descida do valor de *pitch* a partir da palavra QU. No sinal analisado, observa-se que a sílaba inicial possui um valor de *pitch* mais alto que as outras, e em seguida vê-se a curva decrescer até o fim da sentença. Se for levado em consideração o fato de que a sentença é iniciada pela palavra QU, essa sentença comprova a validade do modelo para sentenças de estruturas silábicas distintas da estudada inicialmente.

3.3.1.5 Comando

Para caracterizar essa atitude, as curvas devem ter as mesmas propriedades da atitude Questão QU; entretanto, o conteúdo dessa sentença deve ser imperativo. No sinal observado tem-se um alto valor de *pitch* na sílaba tônica inicial, que correspondia à palavra QU na outra atitude, seguido de uma queda na curva. Essa queda não perdura até o fim da sentença, que apresenta um crescimento no final. No Capítulo 2 pode-se observar que na sentença estudada inicialmente também há um leve crescimento nessa curva, de modo que isso não necessariamente descaracteriza a atitude, e essa atitude. Portanto o modelo pode ser considerado válido, apesar da estrutura silábica.

3.3.1.6 Exclamação QU

No sinal estudado para essa atitude, observa-se um decrescimento do *pitch* a partir da palavra QU, seguindo até o fim da sentença. Esse comportamento era esperado pelo modelo, que é próximo ao da atitude Questão QU. O fator que diferencia essa atitude daquela vista na Questão QU também está presente nesse sinal, pois o valor do *pitch* na sílaba pretônica final é menor do que no sinal que apresenta a atitude

Questão QU. A presença dessas características permite que se considere o modelo dessa atitude como válido, apesar da estrutura silábica distinta.

3.3.2 Atitudes com subida de pitch

3.3.2.1 Questão SN

Seguindo o modelo proposto para essa atitude, o sinal analisado apresenta uma subida na curva de *pitch* na sílaba tônica inicial e um subsequente decrescimento até a sílaba tônica final, onde o sinal volta a crescer. Pelo modelo haveria ainda mais uma queda no sinal situada na sílaba postônica final; entretanto, por se tratar de uma palavra oxítônica, a última palavra da sentença não possui sílaba postônica final. Como a curva corresponde ao previsto no modelo, a menos da diferença causada pela estrutura silábica, pode-se dizer que essa atitude é válida, para diferentes estruturas silábicas.

3.3.2.2 Questão SN Retórica

Para ser classificada como uma sentença que possui as propriedades da atitude Questão SN Retórica, suas curvas características deveriam assemelhar-se às da atitude Questão SN com a diferença de que na sílaba tônica final observar-se-ia um decrescimento do *pitch*, ao passo que na Questão SN essa queda só ocorre na sílaba postônica. Posto que a sentença não apresenta uma sílaba postônica final (já que sua última palavra é oxítônica), não se pode observar tal comportamento na sentença analisada anteriormente. Já na sentença utilizada para essa atitude, vê-se, como esperado, um decrescimento na curva do *pitch* a partir da metade da sílaba tônica final, de modo que se pode considerar o modelo válido para estruturas silábicas distintas.

3.3.2.3 Pedido

De maneira análoga à sentença que expressa a atitude de Questão SN Retórica, essa sentença apresenta uma subida do *pitch* na sílaba tônica inicial, seguida de uma descida, até voltar a subir na tônica final. Como as características para essa atitude não diferem (senão em conteúdo) daquelas da Questão SN Retórica, considera-se

válido o modelo para estruturas silábicas distintas.

3.3.2.4 Questão SN Incrédula

A característica marcante dessa atitude é o vale formado pela subida na tônica inicial, a descida da pretônica final e a nova subida da tônica final. No sinal analisado, observam-se essas características bastante evidentes, de modo que o modelo é válido para estruturas silábicas distintas.

3.3.3 Atitudes com variação de duração

3.3.3.1 Aviso

Conforme predito no modelo dessa atitude, o sinal observado tem um crescimento do valor do *pitch* até a sílaba pretônica final, onde assume seu maior valor, e em seguida apresenta um valor consideravelmente mais baixo na sílaba tônica. As durações e intensidades também apresentam valores correspondentes aos esperados pelo modelo, de forma que este é válido para estruturas silábicas diferentes.

3.3.3.2 Irônica

As curvas de *pitch* dessa atitude são muito parecidas com as da atitude Aviso. A diferenciação entre as duas ocorre na duração, pois nessa o alongamento das sílabas começa a ocorrer na sílaba pretônica final, conforme previsto pelo modelo. Essa característica torna-se, então, uma assinatura característica da atitude e sua presença nesse sinal acaba por validar o modelo para outras estruturas silábicas.

3.3.3.3 Incrédula

A curva de *pitch* desse sinal é levemente decrescente, embora o modelo proposto sugira que deva ser majoritariamente plana, não apresentando picos ou inclinações relevantes. Além disso, há um alongamento das sílabas tônica e pretônica finais, que está de acordo com o modelo. Essa diferença entre o modelo e o sinal, embora existente, não é determinante para dizer que o modelo está errado para outras estruturas; portanto, novo teste será realizado posteriormente.

3.3.3.4 Ênfase Intensiva

Apesar de, no escopo geral, a curva de *pitch* dessa sentença parecer destoar do modelo, essa ainda apresenta um alto valor na sílaba tônica final, que se manteria até a sílaba postônica final caso a palavra não fosse oxítônica, tornando a atitude identificável. Além do *pitch*, pode-se observar outra característica que valida o modelo: o alongamento das sílabas pretônica e tônica finais. Dessa forma, o modelo continua válido para sinais com estruturas silábicas variadas.

3.4 Conclusões

Nesse capítulo, as mesmas 14 atitudes prosódicas vistas no capítulo anterior foram analisadas em sentenças com uma estrutura silábica distinta da usada anteriormente. De forma a validar o modelo dessas atitudes, o comportamento de cada uma na nova sentença foi comparado com aquele previsto pelo modelo. Embora o número de sílabas nas palavras tenha sido mantido ao se escolher as novas sentenças, a modificação da estrutura da sentença causou algumas alterações importantes em algumas atitudes.

Mais especificamente, como a nova estrutura silábica não incluía sílabas pretônicas iniciais ou sílabas postônicas finais, para os modelos baseados nestas unidades ocorrem duas possibilidades: a alteração é ignorada ou é refletida na sílaba tônica mais próxima. Os casos em que isso ocorreu de forma mais clara foram:

- Questão SN: Não há a queda no *pitch* da sílaba postônica final, por se tratar de uma palavra oxítônica e não possuir essa sílaba; nesse caso, essa parte do modelo é descartada.
- Ênfase Intensiva: Não há manutenção do alto valor da sílaba tônica final na sílaba postônica, por se tratar de uma palavra oxítônica, que portanto não possui a referida sílaba; essa alteração é, por isso descartada.

Observadas as alterações mais evidentes, algumas outras mais sutis também se fizeram notar, como valores mais altos ou mais baixos das variáveis analisadas e ainda uma maior presença de atitudes com variação nos valores de intensidade.

Entretanto, de modo geral, essas variações não se mostraram determinantes para invalidar o modelo anterior, pois, dentre outras coisas, a intensidade varia bastante entre as gravações, o que torna seu comportamento pouco previsível.

A partir dos estudos realizados no presente capítulo, os modelos de MORAES [4] para as diferentes atitudes foram generalizados para diferentes estruturas silábicas. No próximo capítulo, vamos formalizar os novos modelos e testá-los para uma nova estrutura silábica distinta das utilizadas até aqui.

Capítulo 4

Aplicação do Modelo e Testes

4.1 Introdução

No capítulo anterior, o modelo das atitudes estudado foi avaliado para uma estrutura diferente daquela para a qual foi proposto inicialmente. No presente capítulo para validar os resultados, o modelo obtido será aplicado a uma sentença de atitude neutra, de modo a impor a atitude desejada.

Na Seção 4.2 será mostrada uma nova estrutura silábica, diferente das já apresentadas, e a seguir será mostrado um exemplo de como aplicar o modelo a uma sentença que contenha esta estrutura.

O procedimento usado para validar os resultados da aplicação do modelo à nova estrutura será descrito na Seção 4.3. Nesse ponto será apresentado o método subjetivo de avaliação utilizado para chegar a esta validação.

Na Seção 4.4, serão discutidos os resultados dos testes e a validade desse modelo para essa estrutura, visando a chegar a uma conclusão sobre a efetividade ou não do método aplicado.

4.2 Aplicação do Modelo Estudado

Nessa seção será feita a aplicação do modelo estudado para as 14 atitudes em novas sentenças contendo uma estrutura silábica até então não explorada, com

o objetivo de validar a generalidade dos modelos obtidos anteriormente em relação à estrutura de sentença utilizada.

4.2.1 A nova estrutura silábica

Para fazer a validação dos modelos de atitude descritos nesse trabalho, serão utilizadas nessa etapa sentenças formadas por duas palavras de três sílabas, sendo a primeira uma oxítônica (portanto tendo a sua última sílaba tônica) e a segunda uma proparoxítônica (nesse caso a sílaba tônica será a primeira da palavra).

De imediato percebe-se que haverá algumas ressalvas a serem feitas a partir do modelo estudado no Capítulo 2, uma vez que a primeira palavra, devido à sua estrutura, não apresenta sílabas postônicas. A sílaba seguinte à tônica, nesse caso, será a tônica final, e a segunda palavra não possuirá uma sílaba pretônica, de modo que esta será a sílaba tônica inicial. Embora à primeira vista isso possa parecer um problema para o método, deve-se lembrar que um caso semelhante ocorre na estrutura silábica estudada no Capítulo 3 deste trabalho.

Conforme estabelecido anteriormente, para contemplar as 14 atitudes propostas por MORAES, será necessário o uso de pelo menos três sentenças diferentes (embora com a mesma estrutura silábica). Essas sentenças diferem entre si no tocante ao seu conteúdo e à presença de palavras-chave para a aplicação das atitudes, como no caso das **palavras QU** anteriormente explicadas.

As sentenças a serem utilizadas são as que se seguem:

- “Olhará a célula.” - Utilizada para representar as atitudes **Pedido**, **Sugestão** e **Comando**. Embora à primeira vista essa sentença pareça possuir um número de sílabas diferente daquele que foi apresentado na descrição acima, sua transcrição fonética mostra a junção da sílaba final da primeira palavra com o artigo que se segue, de forma que ao final, foneticamente falando, existem somente seis sílabas, conforme o planejado.
- “E porque éramos.” - Utilizadas para as atitudes **Questão QU** e **Exclamação QU**. Nesse caso a expressão “E porque” será considerada como palavra QU,

conforme concessão feita anteriormente também para a expressão “E como jogava.”.

- “Liberais éramos.” - Utilizada para as nove atitudes restantes.

4.2.2 Imposição do modelo

Conforme anteriormente apresentado neste trabalho, a atitude de uma dada sentença é determinada pelas características de *pitch*, intensidade e duração das sílabas de uma sentença. De forma a introduzir as características desejadas na sentença de interesse, foram utilizados dois programas computacionais de código aberto: **Prosody Editor**, desenvolvido por LATSCH [1], e **Praat** [6].

A seguir, será mostrado um exemplo de como os programas foram utilizados para a imposição do modelo a uma sentença de atitude neutra, resultando numa sentença manipulada com a atitude desejada:

(i) Inicialmente, utilizando a ferramenta de manipulação prosódica do Prosody Editor, faz-se a alteração dos valores de duração da sentença com atitude neutra. Para isso deve-se alterar a altura das barras de duração correspondentes a cada sílaba, de maneira a fazer com que estas recebam as características explicitadas no modelo.

No caso da atitude **Irônica**, há um aumento da duração da sílaba tônica final frente ao valor apresentado na sentença de atitude neutra. De modo a obter um resultado melhor, as amplitudes das barras de duração foram alteradas para que seus valores máximos e mínimos correspondessem àqueles da sentença de referência (nesse caso, os valores da sentença “Renata jogava.” dita com a atitude irônica). Portanto a duração das sílabas tônicas finais das duas sentenças foi igualada.

Por exemplo, as características da sentença em questão, “Renata jogava.”, podem ser observadas na Figura 4.1, enquanto que na Figura 4.2 observa-se a janela de manipulação prosódica do Prosody Editor em que as características da sentença “Liberais éramos.” foi executada.

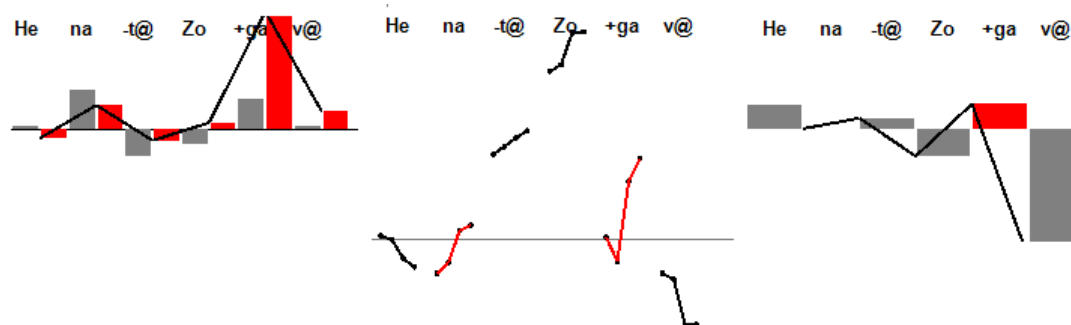


Figura 4.1: Da esquerda para a direita: Gráficos de Intensidade, *Pitch* e duração da sentença “Renata jogava.” com atitude Irônica.

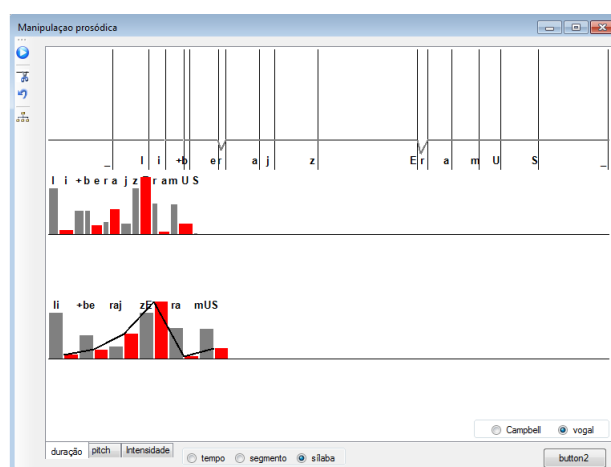


Figura 4.2: Janela de Edição de Duração do Prosody Editor para a sentença “ Liberais éramos.” e atitude **irônica**.

(ii) A seguir, foi alterada a curva de intensidade sonora da sentença de atitude neutra. Novamente nesse caso, em que a atitude de interesse é a **Irônica**, há uma diferença notável na intensidade sonora da sílaba tônica final, de modo que, para manter a coerência com o que foi executado anteriormente, também nessa etapa os valores máximos e mínimos da intensidade foram alterados para aqueles encontrados na sentença de referência “Renata jogava.”. Dessa forma a intensidade sonora da sílaba tônica final das duas sentenças foi igualada. Na Figura 4.3 observa-se a janela de manipulação de intensidade do programa Prosody Editor, enquanto a alteração de intensidade era executada.

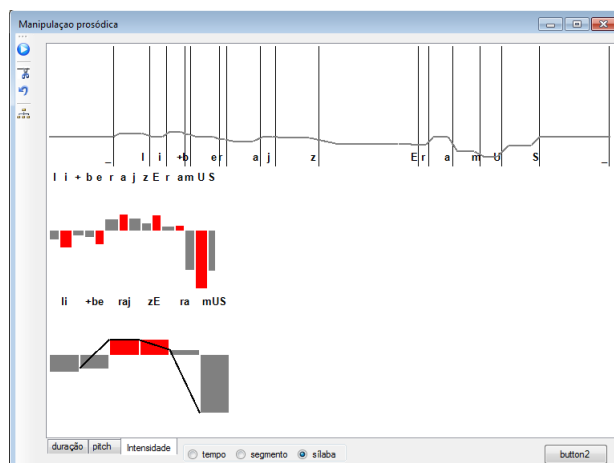


Figura 4.3: Janela de Edição de Intensidade do Prosody Editor para a sentença “Liberais éramos.” e com atitude **irônica**.

(iii) A terceira e última etapa está relacionada à alteração das curvas de *pitch* do sinal com atitude neutra. Essa etapa, assim como as que a precedem, poderia ter sido executada utilizando-se a ferramenta de manipulação de *pitch* do Prosody Editor. Entretanto, essa ferramenta disponibiliza somente um pequeno número de pontos de *pitch* por sílaba, de modo que os valores dessa variável entre os pontos será obtida pela interpolação destes, o que dá pouca liberdade de escolha de valores para o usuário. Para um melhor desempenho foi utilizado o programa Praat, que permite a manipulação direta de um número maior de pontos, e por esse motivo apresenta um resultado superior para esse processo.

No exemplo utilizado, o modelo para a atitude **Irônica** prevê que haja um crescimento no valor do *pitch* no decorrer da sentença, culminando num alto valor dessa variável nas sílabas pretônica e tônica finais, seguido de uma queda no valor do *pitch* até o final da sentença. Conforme esclarecido na Seção 4.2.1 há uma interseção entre as sílabas tônica inicial e pretônica final, o que poderia causar problemas para a definição do comportamento do *pitch* nesse trecho. Entretanto, como se pode notar na Seção 2.4.3.2, o modelo para essa atitude não prevê qualquer comportamento para a sílaba tônica inicial, de forma que essa sílaba pode ser tratada como a pretônica final do modelo.

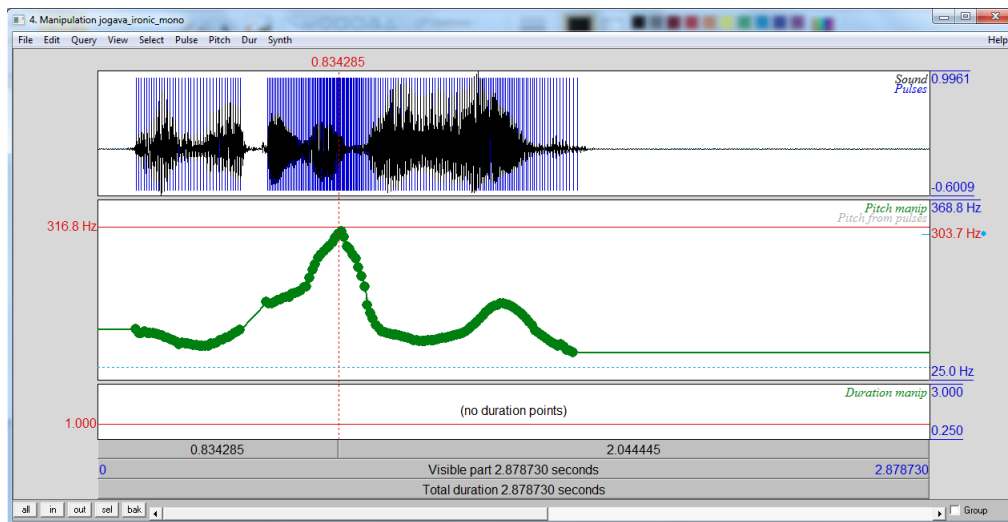


Figura 4.4: Janela de Edição de *Pitch* do Praat para a sentença “Renata jogava.” e com atitude **irônica**.

Segundo o modelo para atitude **Irônica** (presente na seção 2.4.3.2), o *pitch* da sentença deverá ser crescente até as sílabas tônicas e decrescente a partir do final da última delas. Conforme explicitado anteriormente, para obter um melhor resultado, os valores máximos e mínimos de *pitch* da sentença de referência (“Renata jogava.” com atitude Irônica) foram copiados para a sentença manipulada.

As janelas de manipulação de *pitch* do Praat, onde essa etapa foi executada, encontram-se nas Figuras 4.4 e 4.5, em que se observa a correspondência entre os valores mais altos de *pitch*.

O objetivo do presente teste é verificar o grau de similaridade entre as atitudes (entonação) de ambos os sinais. De maneira a quantificar os resultados, cada ouvinte deverá atribuir uma nota inteira entre 1 (um) e 5 (cinco) aos sinais ouvidos, de acordo com a seguinte classificação:

1.Ruim: As atitudes não se parecem nem um pouco;

2.Fraco: Há pouca similaridade entre as duas atitudes;

3.Razoável: Pode-se perceber uma razoável similaridade;

4.Bom: Percebe-se claramente a presença da mesma atitude nos dois sinais;

5.Excelente: As atitudes nos dois sinais são idênticas;

A título de exemplo, a similaridade entre as atitudes dos sinais 1 e 2 pode ser considerada nota 5.

São esperadas algumas distorções no sinal modificado, que devem ser ignoradas pelo ouvinte na avaliação das atitudes. São exemplos dessas distorções: voz metalizada e pequenos picos durante a execução.

Clicando no botão “Sinal Distorcido” podem ser escutados um exemplo de sinal com as distorções esperadas. Durante a execução dos testes, cada sinal poderá ser executado um número qualquer de vezes, à escolha do ouvinte. Para trocar de sinais será necessário atribuir uma nota ao conjunto de sinais atual.

A nota inicial para o sinal é 1 (um). Pode-se alterar essa nota clicando sobre as setas no “slider” da janela.(Para se atribuir a nota 1 deve-se modificar para a nota 2 e posteriormente retornar ao valor 1, dessa forma modificando o valor do “slider” pelo menos uma vez).

Após atribuída a nota aos sinais e pressionado o botão “Próximo Sinal”, não será possível voltar a ouvir o sinal anterior. Ao término das análises, o espaço contendo o número do sinal atual será substituído pelo valor “FIM” e a janela poderá ser fechada. O Teste consiste de 14 atitudes, sendo necessário escutar um total de 28 sinais (comparando-os dois a dois). As atitudes avaliadas não serão identificadas no decorrer do teste para que o ouvinte seja menos influenciado. Os resultados do teste serão armazenados de maneira anônima. O tempo estimado para a conclusão do teste é de 15 minutos.

Para iniciar o teste pressione o botão “Iniciar”. Obrigado por participar!

Na Figura 4.7, pode-se observar a tela em que o ouvinte deverá fazer a avaliação do par de sinais utilizando o *slider* para atribuir uma nota entre 1 (um) e 5 (cinco)

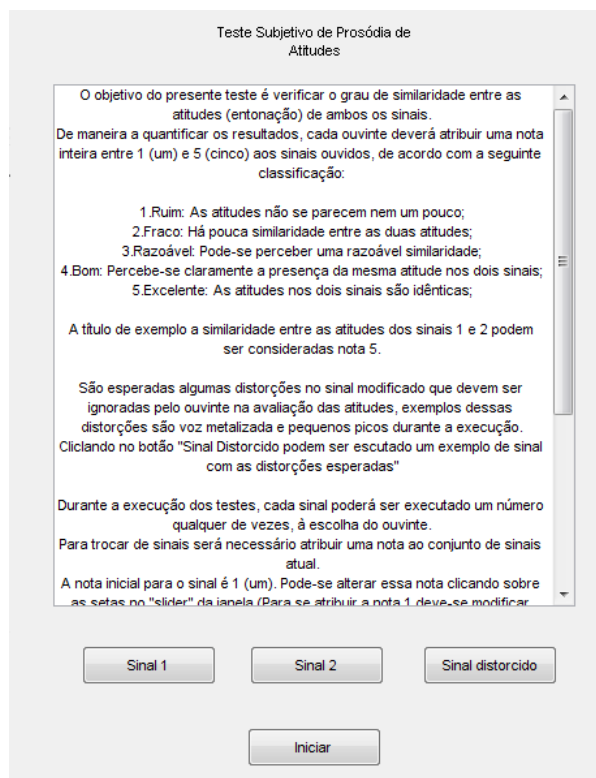


Figura 4.6: Tela inicial do programa de avaliação subjetiva de atitudes onde pode-se observar as instruções para a realização do teste, o texto aqui apresentado está transcrito na página anterior.

a cada par de sinais.

Algumas ressalvas sobre o teste desenvolvido devem ser feitas:

- Conforme descrito no texto acima, o ouvinte deverá atribuir uma nota de 1 a 5 para cada par de sinais. Essa escolha foi feita para que os resultados fossem representados na escala MOS (*Mean Opinion Score*), cujos resultados são bastante conhecidos, e parâmetros de comparação podem ser facilmente traçados.
- Antes do início do teste, ainda na tela inicial, são apresentados ao ouvinte três exemplos de sinal: um par que pode ser considerado merecedor de nota 5, e um sinal que apresenta as distorções mais comuns nos sinais e que para efeito desse teste devem ser desconsideradas. Estas distorções são geradas principalmente na etapa de manipulação de duração do sinal, feita utilizando a ferramenta do Prosody Editor, e não tem relação direta com o modelo a ser avaliado.

Teste de Atitudes

Sinal 1

Referência:

Modificado:

Nota

Figura 4.7: Tela de avaliação do programa de avaliação subjetiva de atitudes.

- Nesse teste o ouvinte não terá acesso ao nome da atitude que está sendo representada pelos sinais que está avaliando, uma vez que este teste não visa a estabelecer se a atitude apresentada corresponde ao conceito que o ouvinte tem sobre a mesma, mas se a atitude presente no sinal de referência foi corretamente refletida no sinal manipulado. Embora possa haver problemas para o usuário entender a atitude a partir do sentido das sentenças essa característica não foi observada nesse teste.

Além de atribuir as notas aos pares de sinais, o ouvinte, ao final do teste, deverá responder a três perguntas: “Qual o par de sinais para o qual encontrou mais dificuldade para atribuir nota?”, “Houve alguma dúvida quanto ao procedimento de avaliação?”, “Houve alguma atitude que não foi capaz de identificar?”. Essas perguntas visam a avaliar a efetividade do teste e aprofundar mais os resultados.

4.4 Resultados e Análises

A Tabela 4.4 mostra os resultados dos testes com as médias referentes a cada atitude e a Figura 4.8 ilustra essas respostas de maneira gráfica.

Tabela 4.1: Médias dos testes subjetivos aplicados

Atitude	1	2	3	4	5	6	7
Média	3,4	3,9	3,0	3,2	3,7	3,1	2,0
Atitude	8	9	10	11	12	13	14
Média	1,8	2,4	2,7	3,0	3,7	3,7	2,3

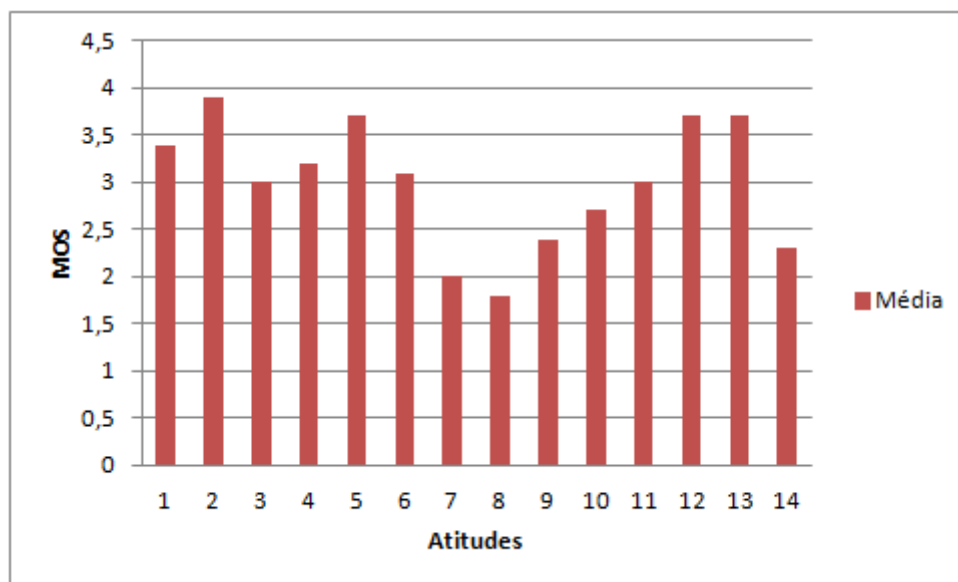


Figura 4.8: Gráficos com as médias dos testes subjetivos aplicados.

Na lista a seguir encontram-se descritas as atitudes representadas por cada par de sinais que aparece na tabela 4.4:

1. Atitude 1: Aviso
2. Atitude 2: Incrédula
3. Atitude 3: Ênfase Intensiva
4. Atitude 4: Irônica

5. Atitude 5: Óbvia
6. Atitude 6: Ênfase Contrastiva
7. Atitude 7: Questão SN
8. Atitude 8: Questão SN Retórica
9. Atitude 9: Questão SN Incrédula
10. Atitude 10: Pedido
11. Atitude 11: Comando
12. Atitude 12: Sugestão
13. Atitude 13: Questão QU
14. Atitude 14: Exclamação QU

Pode-se notar através dos resultados mostrados na Tabela 4.4 que dentre os 14 pares de sinais testados obteve-se um resultado satisfatório para 9 deles, cujas médias ficaram acima de 3,0. Para os cinco sinais restantes obteve-se um resultado pouco satisfatório, uma vez que suas médias ficaram entre 1,0 e 3,0, o que caracteriza, pela definição dada na descrição do teste, uma similaridade inferior a razoável.

As atitudes **Questão SN**, **Questão SN Retórica** e **Pedido** possuem características muito semelhantes e já houve, conforme se observam na Seção 3.4, algumas incoerências do modelo decorrentes da inexistência de algumas sílabas que caracterizam essas atitudes.

A atitude **Questão SN Incrédula** possui como sílaba importante a sílaba pretônica final. Na presente estrutura, silábica essa sílaba se confunde com a tônica inicial, o que pode ser responsável pelo resultado ruim na avaliação dessa atitude.

Observa-se na atitude **Exclamação QU**, que o modelo sugere que o valor do *pitch* da sílaba pretônica deve ser inferior ao da atitude **Questão QU**. Essa característica não é transmitida com precisão pelo método utilizado, embora os valores de *pitch*

sejam copiados de uma sentença de referência. Há ainda o fato de que a sílaba pretônica final é novamente confundida com a sílaba tônica inicial, o que pode ser a característica responsável pelo resultado abaixo do esperado.

Dentre as nove atitudes cujas avaliações obtiveram resultados positivos pode-se destacar o resultado da atitude **Incrédula**, que foi muito próximo ao valor 4.0. É importante notar que no Capítulo 3 a validade do modelo para a atitude **Incrédula** foi posta em cheque, uma vez que havia uma tendência ao decréscimo no sinal analisado. Entretanto percebe-se que o modelo é válido, já que após a imposição da atitude a uma outra sentença o resultado obtido foi um dos melhores.

Um outro fator que deve ser levado em consideração na análise dos resultados é que, embora tenha sido solicitado aos voluntários nos testes subjetivos que os artefatos e ruídos gerados nos sinais pelas imperfeições das ferramentas utilizadas fossem desconsiderados, é possível que o incômodo gerado por esses problemas tenha sido responsável por uma queda nas notas atribuídas e, dessa forma, o resultado apresentado se mostra inferior àquele que se obteria caso esses problemas tivessem sido evitados.

As respostas às perguntas presentes no questionário feito em conjunto com o teste subjetivo podem ser encontradas nas Tabelas 4.2, 4.3 e 4.4 com as estatísticas das respostas referentes às perguntas 1, 2 e 3, respectivamente.

Tabela 4.2: Respostas dadas para a pergunta 1: “Qual o par de sinais para o qual encontrou mais dificuldades para atribuir nota?”.

Sinal	1	2	3	4	5	6	7
Votos	10,0%	5,0%	10,0%	10,0%	15,0%	0,0%	10,0%
Sinal	8	9	10	11	12	13	14
Votos	10,0%	0,00%	15,00%	5,0%	5,0%	5,0%	0,0%

Conforme pode ser observado a partir das respostas à primeira questão (Qual o par de sinais para o qual encontrou mais dificuldade para atribuir nota?), houve um

grande espalhamento nas respostas, significando que a dificuldade na atribuição de notas aos sinais encontra-se majoritariamente na percepção pessoal dos sinais, de forma que não se pode identificar um sinal cujos resultados possam ser explicados diretamente por essas respostas.

Tabela 4.3: Respostas dadas para a pergunta 2: “Houve alguma dúvida quanto ao procedimento de avaliação?”.

SIM	NÃO
5,0%	95,0%

Das respostas obtidas a partir da segunda pergunta do questionário (Houve alguma dúvida quanto ao procedimento de avaliação?), pode-se inferir que o procedimento de avaliação subjetiva dos sinais foi de simples compreensão, não tendo os voluntários apresentado grandes dificuldades na hora de executar o teste, visto que somente um voluntário respondeu com “sim” a essa pergunta.

Tabela 4.4: Respostas dadas para a pergunta 3: “Houve alguma atitude que não foi capaz de identificar?”.

SIM	NÃO
20,0%	80,0%

As respostas apresentadas pelos voluntários à terceira pergunta (Houve alguma atitude que não foi capaz de identificar?) demonstram que foram raros os casos (4 ao todo) em que o mesmo foi incapaz de compreender as atitudes envolvidas no teste, de forma que o seu julgamento da proximidade da atitude do sinal sintetizado àquele usado como referência foi feito de forma consciente. Vale ressaltar que nos casos em que a resposta foi positiva cada voluntário apresentou um sinal diferente no qual mostrou dificuldades, de forma que não há uma atitude específica que tenha se mostrado problemática para os voluntários.

4.5 Conclusões

Nesse capítulo foi apresentada uma nova estrutura silábica para a qual o modelo de atitudes prosódicas deveria ser testado. Observam-se as características principais dessa estrutura, e as possíveis dificuldades que seriam encontradas na imposição do modelo foram apontadas.

Também nesse capítulo foi apresentado o modo através do qual o modelo de atitudes seria imposto a um sinal que inicialmente possuíse atitude neutra, e um exemplo de como isso ocorreria foi mostrado. Em seguida, foi descrito o teste que validaria o experimento e os resultados do mesmo foram analisados.

Observou-se que para a maior parte dos casos (aproximadamente 64%) o método de imposição do modelo de atitudes prosódicas obteve um resultado satisfatório, em que a média dos resultados do teste subjetivo atingiu um valor de pelo menos 3,0 na escala MOS. Isso demonstra que o método, ainda que imperfeito, apresenta resultados positivos.

A Seção 4.4 apresentou possíveis motivos que justificam o resultado aquém do esperado para algumas atitudes, na intenção de permitir que esses problemas possam ser melhor estudados e obtenham resultados mais promissores, com um número maior de atitudes contempladas.

Capítulo 5

Conclusões

5.1 Considerações Finais

Este trabalho teve como objetivo o desenvolvimento e validação de um método de imposição de atitudes prosódicas numa sentença que inicialmente contivesse uma atitude neutra.

A inspiração inicial para esse trabalho foi obtida de sistemas de conversão texto-fala, uma vez que os sinais sintetizados por esses sistemas podem se tornar mais naturais caso sejam capazes de apresentar as atitudes que um locutor humano utilizaria caso fosse ele o autor da sentença.

Inicialmente, com o objetivo de desenvolver o método de imposição de atitudes de prosódia, no Capítulo 2 foi analisado um modelo de atitudes prosódicas que viria a permear todo o trabalho. Esse modelo, proposto inicialmente por MORAES [4] baseia-se em três aspectos do sinal de voz: duração, intensidade e *pitch* para caracterizar cada uma das 14 atitudes: **aviso, incrédula, ênfase intensiva, irônica, óbia, ênfase contrastiva, questão SN, questão SN retórica, questão SN incrédula, pedido, comando, sugestão, questão QU, exclamação QU**. Entretanto esse modelo havia sido trabalhado somente para uma estrutura silábica, aquela presente na sentença “Renata jogava”, ou seja, com 2 palavras de 3 sílabas, ambas paroxítonas. Ainda no Capítulo 2 foram analisados dois métodos que podem ser utilizados para impor uma atitude a um sinal neutro. Entretanto esses métodos são dependentes da estrutura silábica da sentença sendo manipulada. As-

sim, buscou-se uma nova maneira de impor as atitudes às sentenças, sem que as duas possuíssem a mesma estrutura silábica. No Capítulo 3, o modelo de atitudes prosódicas visto anteriormente foi validado de forma a se demonstrar que o mesmo era válido para outras estruturas silábicas fora aquela para a qual foi gerado; para isso, foi utilizada uma nova estrutura com duas palavras de 3 sílabas, a primeira proparoxítona e a segunda oxítona, como em “pífanos tocarão”. Mostrou-se, então, que a utilização do modelo para as aplicações desse trabalho era possível.

Tendo executado a validação do modelo sem que o mesmo estivesse preso à estrutura silábica para a qual estava sendo utilizado, no Capítulo 4 foi apresentada uma maneira através da qual o modelo de atitudes citado anteriormente pode ser imposto sobre uma sentença de atitude neutra. A técnica aqui aplicada requer uma base de dados contendo uma sentença com cada tipo de atitude, embora não seja necessário que essa sentença possua a mesma estrutura silábica que a sentença a ser modificada, dessa forma cumprindo-se o objetivo deste trabalho. Para executar essa etapa de imposição do modelo e testes foi utilizada, ainda, uma nova estrutura silábica, com 2 palavras de 3 sílabas sendo a primeira palavra oxítona e a segunda proparoxítona, como em “liberais éramos”. Ainda no Capítulo 4, foi mostrada a maneira como os resultados do trabalho foram avaliados através de um teste subjetivo realizado com 20 voluntários no qual cada um deveria atribuir uma nota de 1 a 5 para a similaridade entre os sinais sintetizados e sinais de referência. Foram abordados ainda os problemas que o método utilizado apresenta. Como resultado das avaliações, 9 das 14 atitudes inicialmente trabalhadas (**aviso, incrédula, ênfase intensiva, irônica, óbvia, ênfase contrastiva, comando, sugestão, questão QU**) tiveram um resultado considerado aceitável (nota maior que 3.0 na escala MOS), enquanto as 5 atitudes restantes (**questão SN, questão SN retórica, questão SN incrédula, pedido e exclamação QU**) foram avaliadas como inaceitáveis (recebendo notas inferiores a 3,0, na média).

Observa-se ao final do trabalho que, embora nem todas as 14 atitudes inicialmente propostas tenham obtido um resultado satisfatório, o método utilizado alcançou o objetivo proposto, uma vez que sua aplicação independe da estrutura silábica das sentenças envolvidas e permite que as atitudes impostas nas sentenças sejam reco-

nhecidas por ouvintes de maneira adequada. Para as 9 atitudes que obtiveram resultados satisfatórios é possível que se faça a manipulação das características prosódicas de uma sentença de atitude neutra de modo a transformar sua atitude.

5.2 Trabalhos Futuros

O escopo do tema trabalhado é muito maior do que o que foi explorado nesse projeto. Dessa forma é possível que muitas frentes de trabalho sejam abordadas a partir do ponto em que esse trabalho se encerrou. Dentre os caminhos que podem ser trilhados a partir desse ponto destacam-se:

- Explorar as exceções ao modelo, de maneira que as atitudes que não obtiveram bons resultados possam ser melhor caracterizadas e dessa forma a manipulação de sentenças com o objetivo de impor essas atitudes obtenha melhores resultados;
- Incorporar o método desenvolvido a sistemas de conversão texto-fala (como por exemplo a ferramenta contida no sistema SASPRO), de modo que estes possam fazer a transformação das atitudes sem que as duas sentenças possuíssem a mesma estrutura silábica e sem que os mesmos precisem conter um banco de atitudes praticamente infinito para cobrir todas as combinações de atitudes e estruturas silábicas da língua portuguesa;
- Ampliar o número de atitudes previstas pelo modelo, fazendo assim com que o modelo possa cobrir um número maior de emoções humanas e permitindo que um sistema TTS possa expressar as suas sentenças de maneira mais natural;

Referências Bibliográficas

- [1] LATSCH, V. L., *Desenvolvimento de um sistema de conversão texto-fala com modelagem de prosódia*. Tese de doutorado, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Junho 2011.
- [2] DUTOIT, T., STYLIANOU, Y., *The Oxford Handbook of Computational Linguistics*, chapter 17 - Text-to-Speech Synthesis, New York, Oxford University Press, pp. 323–334, 2005.
- [3] SIMÕES, F. O., *Implementação de um Sistema de Conversão Texto-Fala para o Português do Brasil*. Tese de doutorado, Unicamp, São Paulo, Maio 1999.
- [4] DE MORAES, J. A., “The Pitch Accents in Brazilian Portuguese: analysis by synthesis”, *Proceedings of the Fourth Conference on Speech Prosody*, pp. 389–398, Maio 2008.
- [5] BAILLY, G., HOLM, B., “SFC: A trainable prosodic model”, *Speech Communication*, v. 46, pp. 348–364, Abril 2005.
- [6] BOERSMA, P., WEENINK, D., “Praat: doing phonetics by computer. Version 5.3.22”, <http://www.praat.org/>, 2012, (Acesso em 23 Julho 2012) [Programa Computacional].